

Climate and Chronology: Environmental variability and the archaeology of Marine Isotope Stage 3



Monty Ochocki
Harris Manchester College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2023

Acknowledgements

I would like to thank my supervisors, Tom Higham and Christopher Bronk Ramsey, for their advice, enthusiasm, and expertise given throughout my DPhil. I would also like to thank Edward Armstrong and Paul Valdes at the University of Bristol for providing access to their data, for their technical guidance, and for their collaborative approach to research. Many thanks to Katerina Douka and researchers at the University of Siena for collecting and allowing access to samples from Paglicci, and to Rachel Wood for her support and guidance throughout the radiocarbon experiments. I am also immensely grateful to all the people at the ORAU that have helped so much and in so many ways during my DPhil. Thank you for your help, friendliness, and patience.

I am also very grateful to NERC for financial support of my research, and for the opportunity to be a student of the Doctoral Training Programme in Environmental Research. Thank you to Victoria Forth, and to the friends I've made through the DTP.

I would like to thank my friends in Oxford and elsewhere for all of the good times shared over the past five years. I am also very grateful to have had such understanding and supportive housemates, thank you for your generosity and guidance. Finally, I would like to thank my family for their encouragement and support, especially Kate.

Monty Ochocki
Harris Manchester College, Oxford
6 October 2023

Abstract

The relationship between large-scale palaeoclimatic variability and changes in human behaviour, technology, and distribution is frequently discussed in European Palaeolithic archaeology. However, making confident and causal links between these processes requires addressing some key chronological uncertainties and broader difficulties in palaeoenvironmental reconstruction. This thesis explores and illuminates some of these challenges, by focusing on the themes of *Climate* and *Chronology*.

During Marine Isotope Stage 3 (MIS 3; 60 to 30 thousand years before present), dramatic and abrupt temperature oscillations are evidenced by Greenland ice core records, implying significant changes to the earth's climate system. These changes are likely to have had environmental and ecological impact across the northern hemisphere. However, the spatial footprint of these changes over Europe is unclear, with many palaeoenvironmental records showing discrepancies with the Greenland records during MIS 3, as well as with each other. A key challenge in this field is chronological. Obtaining precise and accurate radiocarbon determinations for both the archaeological and palaeoenvironmental records of MIS 3 can be very difficult, meaning that addressing and ascribing causality across their respective sequence is challenging.

Reconstructing the impact of these palaeoclimatic change demands the synthesis of a broad range of data types, as well as their associated uncertainties. This thesis addresses this challenge, focusing on the chronologies of palaeoenvironmental and archaeological changes during MIS 3. The research explores multiple methods to improve palaeoenvironmental reconstructions and their chronologies, employing General Circulation Models, Machine Learning and Bayesian statistics, and signal processing algorithms. Alongside computational approaches to these problems, laboratory-based research contributes experimental pretreatment protocols and updated estimates for the timings of archaeological changes in Palaeolithic Europe.

The various chapters of this research show that claims for a causal relationship between large-scale environmental change and archaeological shifts during the Late Pleistocene cannot be supported by current evidence. This is not to say there was no relationship between these processes. Rather, current chronologies and their associated uncertainties are not appropriate to answer the questions being asked of the data.

Contents

List of Figures	ix
List of Tables	xx
List of Abbreviations	xxiii
1 Introduction	1
Background	3
Research Aim and Objectives	7
Thesis Structure and Chapter Outline	8
2 Statistical Downscaling of a GCM Simulation Over Europe During MIS 3	12
Abstract	12
2.1 Introduction	14
2.2 Materials and methods	19
2.2.1 Study Region	19
2.2.2 General Circulation Model	19
2.2.3 Statistical Downscaling Scheme	20
2.2.3.1 Generalised Additive Models	22
2.2.3.2 Response Variables	23
2.2.3.3 Predictor Variables	23
2.2.3.4 Spatial Cross-validation and Feature Selection	26
2.2.4 Köppen-Geiger Scheme	29
2.2.5 MIS 3 Downscaling and Validation	31
2.3 Results	34
2.3.1 Feature Selection	34
2.3.2 GAM Performance	35
2.3.3 MIS 3 Downscaling	37
2.3.4 MIS 3 Validation	41

2.4	Discussion	45
2.4.1	GAM Performance and Spatial Cross-Validation	48
2.4.2	MIS 3 Downscaling and Validation	51
2.4.3	Relevance for the Archaeology of the Middle to Upper Palaeolithic Transition	55
2.5	Conclusion	58
3	Synchronised Climate Proxies for Marine Isotope Stage 3 by Dynamic Time Warping	59
	Abstract	59
3.1	Introduction	61
3.2	Materials and Methods	65
3.2.1	DTW Algorithm	65
3.2.2	DTW for Palaeoenvironmental Synchronisation and the MVM Algorithm	68
3.2.3	Palaeoenvironmental Records	71
3.2.3.1	Downscaled Reconstructions	72
3.2.3.2	Tenaghi Philippon	73
3.2.3.3	Monticchio	74
3.2.3.4	Ioannina	75
3.3	Results	78
3.3.1	Tenaghi Philippon	79
3.3.2	Monticchio	81
3.3.3	Ioannina	83
3.4	Discussion	85
3.5	Conclusion	91
4	Dendrochronological Applications of Dynamic Time Warping	92
	Abstract	92
4.1	Introduction	94
4.2	Materials and Methods	98
4.2.1	Towai Farm Dataset	98
4.2.2	Sampling, Detrending, and Missing and Added Rings	101
4.2.3	Alignment by Pearson Correlation and t -Test	102
4.2.4	Alignment by Dynamic Time Warping	104
4.2.4.1	Including Correlation in the DTW Distance Matrix (D)	105

4.2.4.2	Constraining Alignments to Linear Solutions	105
4.2.4.3	DTW and Adjusted t -Test Statistic	106
4.2.4.4	DTW Distance and Sliding Window Approach	107
4.2.5	Assessing Accuracy	108
4.3	Results	109
4.4	Discussion	115
4.4.1	Distribution of Samples and Alignments	116
4.4.2	Parameter Sensitivity and Variable Sample Lengths	120
4.4.3	Accuracy Metrics	123
4.4.4	Directions and Recommendations for Future Research	125
4.5	Conclusion	128
5	Acid-Base-Acid with Stepped Combustion (ABA-SC) as a Pre-treatment Method for Radiocarbon Dating of Charcoal	129
	Abstract	129
5.1	Introduction	131
5.1.1	Contamination in Charcoal	134
5.1.2	Pretreatment Methods for Radiocarbon Dating of Charcoal	137
5.1.3	A Review of Previous Studies Comparing ABA-SC and ABOx-SC	141
5.1.4	Summary	146
5.2	Materials and Methods	148
5.2.1	Samples	148
5.2.1.1	Grotta di Fumane	148
5.2.1.2	Maninjau Charcoal	150
5.2.2	ABA	151
5.2.3	Precombustion	152
5.2.4	Graphitisation and AMS Measurement	154
5.2.5	Calculating the Proportional Effect of Stepped Combustion in Comparison to ABOx-SC	155
5.2.6	Duplicate ABA Measurements	156
5.3	Results	157
5.3.1	Radiocarbon Measurements	157
5.3.2	Proportional Effect of Precombustion on Contaminant Removal	161
5.3.3	Duplicate ABA Measurements and the Effect of Long-term Storage at the ORAU	161

5.4	Discussion	164
5.4.1	Radiocarbon Measurements and Contamination Removal by Precombustion	164
5.4.2	%C and $\delta^{13}\text{C}$ Measurements for ABA-SC	166
5.4.3	Long-term Storage of Excess Material at the ORAU	169
5.5	Conclusion	171
6	Acid-Bleach with Stepped Combustion (ABI-SC) as a Pretreat- ment Method for Radiocarbon Dating of Charcoal	172
	Abstract	172
6.1	Introduction	174
6.2	Materials and Methods	178
6.2.1	Samples	178
6.2.2	Pretreatment Chemistry and the ABI-SC Method	179
6.3	Results	181
6.4	Discussion	186
6.5	Conclusion	190
7	Discussion of Experiments in Radiocarbon Dating	191
	Abstract	191
7.1	Summary of Experiments in Radiocarbon Dating of Charcoal	192
7.2	Samples Pretreated by Both ABA-SC and ABI-SC	193
7.3	Implications of Experimental Results for Charcoal Pretreatment	196
7.4	The Impact of Sample Selection on Experimental Pretreatment Meth- ods and Using ABOx-SC as a Benchmark	200
7.5	Challenges and Recommendations for Future Research	203
8	An Updated Radiocarbon Chronology for the Upper Palaeolithic Sequence at Grotta Paglicci, Italy	209
	Abstract	209
8.1	Introduction	211
8.2	Grotta Paglicci	216
8.2.1	Archaeological Sequence	216
8.2.2	Previous Radiocarbon Measurements and the 23C2 Tephra	220
8.2.3	Palaeoenvironmental Context	224
8.2.4	Summary	228
8.3	Materials and Methods	229

8.3.1	Radiocarbon Dating	229
8.3.1.1	Sample Selection	229
8.3.1.2	Sample Pretreatment and AMS Measurement	229
8.3.2	Bayesian Modelling	230
8.3.2.1	Full Model	231
8.3.2.2	Partial Model	231
8.3.2.3	Outlier Detection	231
8.3.3	Dating the 23C2 Tephra	232
8.3.4	Comparison to Palaeoenvironmental Proxies	233
8.4	Results	235
8.4.1	Full Model	238
8.4.2	Partial Model	238
8.4.3	Dating of the 23C2 Tephra	238
8.4.4	Comparison to Environmental Records	241
8.5	Discussion	243
8.5.1	Previous Chronologies at Paglicci	243
8.5.2	Bayesian Models	246
8.5.3	Estimates for the Emergence of the Aurignacian and Gravettian at Paglicci	249
8.5.4	Relationship Between Environmental Variability and Archaeological Changes at Paglicci	251
8.6	Conclusion	255
Conclusion		256
References		262
Appendices		
A Dynamic Time Warping Function		298
B Correlation-based Dynamic Time Warping Functions		300
C OxCal Code for Full Bayesian Model for Grotta Paglicci		302
D OxCal Code for Partial Bayesian Model for Grotta Paglicci		307

List of Figures

1.1	An example calibrated radiocarbon measurement of 40000 ± 500 ^{14}C years (blue shading), calibrating to 44207-42654 calBP at the 95.4% confidence interval (blue line). The calibrated date is plotted alongside the Greenlandic temperature reconstruction (black line, Kindler et al. 2014) and its associated chronological uncertainty (grey shading; Rasmussen et al., 2014). Considering the uncertainties of the radiocarbon date and Greenland record, it is not clear whether the date relates to a warm, cold, or transitional period of MIS 3.	6
1.2	Chapter structure.	9
1.3	Locations of palaeoenvironmental records and archaeological sites considered in this thesis.	10
2.1	Response variables over the study region. Panel A shows mean monthly temperature observations, and Panel B total annual precipitation (New et al. 2002). Panels C and D show probability density functions for mean monthly temperature and precipitation variables, respectively, over all months and gridcells. Temperature shows an approximately Gaussian distribution, whilst precipitation shows a Gamma distribution.	24
2.2	Sea-level curve used to adjust the SRTM-30 Plus DEM, reconstructed from the Red Sea sediment core LC21 (Grant et al. 2012).	25
2.3	Predictor variables used to build GAMs. For predictors that vary by month (Panels A-D, G, and H), surfaces for July are shown. . .	27
2.4	Locations of palaeoenvironmental records used to validate GAM predictions.	32
2.5	Records used to validate downscaled predictions for MIS 3. References for validation records are provided in Table 2.3.	33
2.6	Quantile-quantile plots and scatter plots of residuals against fitted values for Temperature (Panels A and B, respectively) and Precipitation (Panels C and D, respectively), for OSE (blue dots) and BIC (red dots) feature selection methods.	36

2.7	Monthly distribution of root mean squared error (RMSE) for Bayesian-Information-Criteria (BIC, red) and one standard error rule (OSE, blue) feature selection methods. Vertical error bars indicate one standard deviation above and below the mean, calculated across spatial folds.	37
2.8	Spatial distribution of residuals (observations minus fitted values) over study region for January. Panel A shows temperature residuals, panel B precipitation. Polygons show spatial folds used in cross-validation.	38
2.9	GAM link functions fitted to predictors for temperature.	39
2.10	GAM link functions fitted to predictors for precipitation.	40
2.11	Response vs. fitted values and histograms of residuals for temperature (Panels A and C, respectively) and precipitation (Panels B and D).	41
2.12	Mean annual temperature and annual precipitation predicted by GAMs for the study region over MIS 3. Panel A shows mean annual temperature (red line) with minimum and maximum temperatures above and below, respectively (grey lines). Panel B shows total annual precipitation averaged over the study region. Panel C shows the standard deviation of monthly temperatures, giving some indication of seasonality. Panel D shows the coefficient of variation of monthly total precipitation. Panel E shows NGRIP temperature reconstructions for comparison (Kindler et al. 2014).	42
2.13	Spatial trends over MIS 3. Panels A and B show the standard deviation of mean annual temperatures over MIS 3 and the log of the standard deviation of total annual precipitation over MIS 3, respectively. Greater variability is shown for both temperature and precipitation over western areas of the study region. Panel C shows the correlation coefficient of mean annual temperatures and total annual precipitation for MIS 3. Panel D shows the mean difference in annual temperatures between GAM predictions and the delta method, suggesting overall colder conditions predicted by the delta method.	43

2.14	Köppen-Geiger classification over MIS 3, based on GAM predictions. Panel A shows NGRIP temperature reconstructions (Kindler et al., 2014). Panel B shows the percentage cover for each climate classification over MIS 3. Expansion and contraction of Arid (B), Temperate (C), and Cold (D) categories are particularly apparent, corresponding to DO cycles as shown in Panel A. Further description of the Köppen-Geiger classification scheme and its categories is given in Table 2.2, following Willmes et al. (2017).	44
2.15	Temperature and precipitation anomalies (CRU minus GAM predictions) for 43.7 kyr BP (Panels A and C) and 46.7 kyr BP (Panels B and D). Panels E and F show Köppen-Geiger classifications over the study region for 43.7 and 46.7 kyr BP, respectively.	45
2.16	Comparisons of palaeoenvironmental proxies (green) and temperature (red) and precipitation (blue) predictions produced by GCM downscaling at their respective locations. Correlation coefficients are shown at the top right corner of each panel, for temperature (red) and precipitation (blue).	46
2.17	Density plots for values of GCM T and P (Panels A and B, respectively), for training data (0 ky BP, red) and two example time steps, GI 12 (46.7 ky BP, blue) and GS 12 (43.7 ky BP, green). T for GS 12 is distributed over much colder temperatures than training data, requiring extrapolation of the GAM function.	52
3.1	Example age model for the Ioannina I-284 pollen record, used for validation in Chapter 2. Panel A shows % Temperate Forest pollen for the Ioannina core (green line). Panel B shows the existing age model for Ioannina I-284 between 20 and 65 kyr BP (Sánchez Goñi et al., 2017). The existing chronology is constructed by linear interpolation between a single radiocarbon date (blue line) around 25 ky calBP and a tie point (red dot) at 59 kyr BP, resulting in an unrealistic linear age model for MIS 3. As a result, the pollen signal and its timing are uncertain, and may not be appropriate as validation for downscaled models in Chapter 2.	63

3.2	An example alignment problem and solution by Symmetric DTW. Panel A shows two signals: a reference sine wave (blue line); and a query (red line), to be aligned to the reference. Panel B shows the query realigned after Symmetric DTW. Panel C shows the distance matrix D (Equation (3.1)), here given by the Euclidean distance between the query and reference at all possible offsets. Panel D shows the cumulative cost matrix C constructed by the Symmetric step pattern (Equation (3.2)). The warping path, drawn by a least cost path across C and representing the realignment of the query indices to those of the reference, is overlain on Panels C and D in white.	66
3.3	Comparison of step patterns in DTW: Symmetric (left, Equation (3.2)); Asymmetric (middle left, Equation (3.3)); Symmetric 2 (middle right, Equation (3.4)); MVM with elasticity (L) = 3 (right, Equation (3.5)). Black dots represent candidates for alignment to black circles. Numbers next to lines connecting dots and circles are multipliers for weighting alignment.	68
3.4	The example alignment problem shown in Figure 3.2 and its solution by DTW with the Minimal Variance Matching (MVM) step pattern. Panel A shows the query and reference signals. Panel B shows the realigned query. Panel C shows the distance matrix D , as for Figure 3.2. Panel D shows the cumulative cost matrix C constructed by the MVM step pattern with $L = 3$ (Equation (3.5)). In Panel C and D, the warping path is overlain in white. Under MVM, the warping path cannot be vertical or horizontal, meaning that no two indices of the query can be aligned to a single index of the reference, and vice-versa. Implicitly, areas of the cost matrix are undefined under MVM.	70
3.5	Locations of the pollen records used in this study.	72
3.6	The Tenaghi Philippon pollen record and its existing chronology. Panel A shows % Arboreal pollen from the Tenaghi Philippon sediment core (green line), compared to downscaled temperature (red line) and precipitation (blue line) predictions. Vertical lines in Panel B shows locations of chronological information used to build the age model, by radiocarbon measurements (dark green), manual tuning ("Event" layers; orange), and tephra layers (purple).	74
3.7	% Temperate Forest pollen from the Monticchio sediment core (green line) based on its existing age model, compared to downscaled temperature (red line) and precipitation (blue line) predictions. As the Monticchio age model is based on varve counting, no additional chronological information is shown.	76

3.8	The Ioannina I-284 pollen record. Panel A shows % Temperate Forest pollen from the Ioannina sediment core (green line) based on its existing age model, compared to downscaled temperature (red line) and precipitation (blue line) predictions. Panel B shows the location of the single tuning point at 59 kyr BP used to build the Ioannina chronology.	77
3.9	Comparison between original age models (black lines) and age models produced by DTW alignment to downscaled temperature (red lines) and downscaled precipitation (blue lines), at Tenaghi Philippon (Panel A), Monticchio (Panel B), and Ioannina (Panel C). Grey shading before 60 kyr BP and after 30 kyr BP show regions not considered for alignment.	79
3.10	Realigned signals for Tenaghi Philippon. Panels A and B shows the cumulative cost matrix (C) with warping path for alignment by DTW to downscaled temperature and precipitation, respectively. Panels C and D show the realignment of the Tenaghi Philippon pollen record (red and blue lines, respectively, left y-axes) to the downscaled temperature and precipitation predictions (grey lines, right y-axes). Panel E shows the pollen signal plotted to its original chronology.	80
3.11	Realigned signals for Monticchio. Panels A and B shows the cumulative cost matrix (C) with warping path for alignment by DTW to downscaled temperature and precipitation, respectively. Panels C and D show the realignment of the Tenaghi Philippon pollen record (red and blue lines, respectively, left y-axes) to the downscaled temperature and precipitation predictions (grey lines, right y-axes). Panel E shows the pollen signal plotted to its original chronology. .	82
3.12	Realigned signals for Ioannina. Panels A and B shows the cumulative cost matrix (C) with warping path for alignment by DTW to downscaled temperature and precipitation, respectively. Panels C and D show the realignment of the Tenaghi Philippon pollen record (red and blue lines, respectively, left y-axes) to the downscaled temperature and precipitation predictions (grey lines, right y-axes). Panel E shows the pollen signal plotted to its original chronology. .	84
3.13	An example of DTW alignment for Ioannina where parameters "open.end" and "open.begin" are used. When the resolution of the query is relatively low, and without constraining the start and end of the query to those of the reference, the query signal is likely to be compressed to subsections of the reference.	87

4.1	Location of Towai Farm (red triangle) in New Zealand.	99
4.2	Towai Farm tree ring series and resulting master sequence (Palmer et. al., 2016). Panel A shows the individual sequences from the Towai series in their original alignment, from which samples of varying length were extracted and cross dated. Panel B shows the resulting Towai master sequence ring width indices (RWI), produced by log transformation, spline detrending and averaging across sequences (frequency response = 50%; wavelength = 50-years).	100
4.3	"Signal-free" detrended Towai master sequence and resulting ring width indices (RWI; black line; left y-axis), with 50-year spline overlain (frequency response = 20%; red line), and sample depth (light blue shading; right y-axis).	103
4.4	An example alignment problem and it's solution by DTW and Pearson correlation. In Panel A, the reference vector (black line) and query (red line) are shown, with the query positioned at the correct index (i.e., the solution). Panel B shows the scores of sliding window alignment by DTW, the maximum of which gives the solution. Panel C is as for Panel B, but for Pearson correlation.	107
4.5	Boxplots summarising the accuracy of alignments by DTW (red) and Pearson correlation (blue) at varying sequence lengths between 20 and 100 (x-axis), for all numbers of rings added or removed. . . .	109
4.6	Sampling distribution and frequencies of correct and incorrect matches over the Towai series. Panel A shows the Towai master sequence after spline detrending, using all sequences. Panel B shows the Towai series depth (black line) and sampling distribution over the series' indices, coloured by sequence length (bin width = 10). Panel C shows the distribution of correct sequence matches for DTW (dark grey) and Pearson correlation (light grey) alignment methods. Panel D shows the distribution of incorrectly matched indices for DTW and Pearson correlation, showing an alignment bias for Pearson correlation around index 1100.	111
4.7	Accuracy of alignments for varying sequence lengths by DTW (top row) and by Pearson correlation (bottom row). Values in cells show the proportion of correctly matched samples for rings added and rings removed. A correct alignment is defined as being within 5 indices of the original sample location. Rings added are shown on horizontal axes (from 0 to 5), rings removed are shown on vertical axes.	112

4.8	Paired DTW (α) and Pearson correlation (r) scores for correct alignments. Panel A shows the relationship between α and r for cases where both methods found the correct solution, coloured by sequence length. A positive relationship between α and r suggests that for cases where Pearson correlation is high, DTW scores are also typically high. Panel B shows the same paired scores after adjusting for sample length by t -Test. Densities of α , r , and t -Test scores are shown along the upper and right axes of both panels, showing less clear separation after t -Test adjustment.	113
4.9	Paired DTW (α) and Pearson correlation (r) scores for incorrect alignments. Panel A shows the relationship between α and r for cases where both methods did not find the correct solution, coloured by sequence length, as for Figure 3.8. Panel B shows the same paired scores after adjusting for sample length by t -Test. Densities of α , r , and t -Test scores are shown along the upper and right axes of both panels, showing less clear separation after t -Test adjustment.	114
4.10	The Towai master sequence from indices 1120 to 1240, where some bias is apparent for alignment by Pearson correlation. Panel A shows the Towai master sequence (black line, left y-axis), series depth (blue shading, right y-axis), and highlights the region between indices 1120 and 1240 (red vertical lines and shaded region). Panel B shows indices 1120 to 1240 of the Towai master sequence. Incorrect alignments by Pearson correlation are shown in red ($N = 761$), with the master sequence overlain in black.	117
4.11	Two-dimensional density plot of incorrect alignments by Pearson correlation, showing the original Towai indices (x-axis) and the alignment indices (i.e., where samples were incorrectly aligned; y-axis). Some bias to alignment is apparent around index 1200, with higher density in this region (y-axis).	118
4.12	Two-dimensional density plot for incorrect alignments by DTW, showing original Towai indices (x-axis) and the alignment index (i.e., where samples were aligned; y-axis).	119
4.13	The relationship between interseries correlation (x-axis) and % accuracy (y-axis) for each sequence in the Towai series (e.g., TOW101A), for DTW (blue) and Pearson correlation (red), with linear regression and R^2 values. Generally, both alignment methods were more accurate where interseries correlation was higher.	121
4.14	Distribution of differences between original index (i.e., the solution) and alignments by DTW and Pearson correlation, for correct matches (alignment falling within the accuracy threshold of 10 years).	124

4.15	Densities of t -Test scores for correct (blue) and incorrect (yellow) alignments, for Pearson correlation (Panel A) and DTW (Panel B). The t -Test threshold of 3.5 is shown by a dashed line.	125
5.1	The effect of modern and fossil contamination in radiocarbon measurement. As the actual ^{14}C age of the sample increases, modern contamination has an increasingly significant effect on the measured ^{14}C age, whilst for the inclusion of fossil (" ^{14}C -dead", dotted line) contamination, the effect is linear.	132
5.2	Location of Grotta di Fumane (red triangle).	149
5.3	Stratigraphic diagram of relevant layers at Grotta di Fumane, Italy. Original image from Higham et al. (2009). Key: 1: denotes rendzina, upper soil; 2 slope deposits with boulders; 3 living floors, with high concentrations of organic matter or charcoal; 4 loess and sandy loess; 5 CaCO_3 cemented layers; 5 sandy deposits; 6 unweathered and weathered bedrock.	151
5.4	Schematic diagram of the ABA, ABOx-SC, and ABA-SC methods, as practiced at the ORAU.	154
5.5	Summary of new ABA-SC radiocarbon determinations from Fumane, with ABA and ABOx-SC dates for comparison. Panel A shows radiocarbon ages in ^{14}C years for each sample for ABA, ABA-SC, and ABOx-SC pretreatment methods. Panel B shows stable carbon isotope ratios for the samples and methods. Panel C shows the proportional effect of precombustion on ABA samples for contaminant removal, relative to the inferred amount removed by ABOx-SC (the higher the effect, the more contamination was removed, expressed as a proportion of the amount removed by ABOx-SC). Red dashed line and grey shading in Panel C show the mean and standard deviation across samples for R	158
5.6	Boxplots showing $\%C$ measurements for ABA, ABA-SC, and ABOx-SC. Numbers over boxplots indicate the number of samples summarised for each pretreatment.	161
5.7	Comparison of duplicate ABA radiocarbon and stable isotope measurements for Fumane samples made at the ORAU since 1996. Panel A shows radiocarbon measurements for each sample in radiocarbon years. All measurements made in 2023 (purple) were made for this experiment. Panel B shows stable carbon isotope ratios.	163

5.8	Paired χ^2 values for original and updated ABA radiocarbon dates (y-axis) against the number of years between measurements (Years in Storage, x-axis). Dashed horizontal line shows the value above which χ^2 values are significant (3.84, df = 1, N = 2).	169
6.1	Schematic diagram of the ABA, ABOx-SC, AOx-SC, ABl, and ABl-SC pretreatment methods. ABl and ABl-SC are experimental pretreatment methods, whilst ABA, ABOx-Sc, and AOx-SC are routinely practiced at the ORAU (Brock et al., 2010).	181
6.2	Summary of new ABl and ABl-SC radiocarbon determinations from Fumane, with ABA and ABOx-SC dates for comparison. Panel A shows radiocarbon ages for each sample for ABA, ABl, ABl-SC, and ABOx-SC pretreatment methods. Panel B shows stable carbon isotope ratios for the samples and methods. Panel C shows the proportional effect of ABl or ABl-SC over ABA, relative to the improvement over ABA by ABOx-SC (the higher the effect, the more contamination was removed, expressed as a proportion of the amount removed by ABOx-SC). Red dashed line and grey shading in Panel C show the mean and standard deviation across samples for R	184
6.3	Boxplots showing %C measurements for ABA, ABl, ABl-SC, and ABOx-SC pretreatment methods. Numbers over boxplots indicate the number of samples summarised for each pretreatment.	185
7.1	Three samples from Fumane dated by ABA-SC, ABl, and ABl-SC, compared to their ABA and ABOx-SC dates and $\delta^{13}\text{C}$ measurements. Panel A shows radiocarbon determinations for the samples, in ^{14}C years. Panel B shows $\delta^{13}\text{C}$ measurements. Panel C shows the improvement of each pretreatment method over ABA, expressed as a proportion of the improvement by ABOx-SC (Equation (5.1), Chapter 5).	194
7.2	All available dates made at the ORAU for samples from Fumane, dated by ABA-SC, ABl, and ABl-SC methods (blue, purple, and orange, respectively). These include multiple ABA dates for each sample (red), ABOx without precombustion (grey), ABOx with CuO precombustion (green), ABOx with O_2 (pink) precombustion, and plasma ashed and HCl pretreatment (brown). Panel A shows radiocarbon determinations for each sample. Panel B shows $\delta^{13}\text{C}$ measurements for each sample. Considerable spread is particularly noticeable for ABA dates.	197

8.1	$\delta^{18}\text{O}$ record from NGRIP (black) with its chronological uncertainty in grey (Rasmussen et al., 2014), and approximate timings for the emergence and duration of Middle and Upper (Aurignacian and Gravettian) Palaeolithic industries in Europe (Douka et al., 2020; Higham et al., 2014).	213
8.2	Location of Grotta Paglicci in Italy (top left), its location on the Gargano Promontory (bottom left), and the stratigraphic diagram of layers of the interior cave sequence and major archaeological phases at the site (right, redrawn from Boschini et al., 2018).	217
8.3	Previous radiocarbon determinations for Grotta Paglicci individually calibrated using IntCal20 in OxCal. Dates are labelled by their site layer and sublayer (e.g. 3A), as well as their lab code and number where known (F = Florence; Ly = Lyon; GrN = Groningen; UtC = Utrecht; R = Rome).	222
8.4	Environmental records for MIS 3. The NGRIP record (black line) and associated chronological uncertainty (grey shading) is shown above, converted to the IntCal time scale (Rasmussen et al., 2014). % arboreal pollen at Tenaghi Philippon (red; Müller et al., 2011), and % forest pollen at Ioannina (blue; Tzedakis et al., 2002) and Monticchio (green; Allen et al., 1999) are shown below. Interglacial periods defined by Rasmussen et al., (2014) are overlain with grey shading.	226
8.5	Unmodelled and calibrated radiocarbon measurements from Grotta Paglicci. Labels indicate sample layer (e.g. 18B2), OxA number, and pretreatment method (e.g. ABA). OxA-27737 from Layer 21A1 was suspected to be contaminated given its poor agreement with adjacent dates, therefore was redated using AOx-SC. The AOx-SC measurement (OxA-P33365) is more consistent with adjacent dates and the site stratigraphy.	237
8.6	Full Bayesian model, built using new radiocarbon measurements and existing published dates from Grotta Paglicci. Detected outliers are highlighted in red.	239
8.7	Partial Bayesian model, using only radiocarbon dates produced in this study. Detected outliers are highlighted in red.	240
8.8	Age estimates for the 23C2 tephra layer, extracted from Full (dark grey) and Partial (light grey) models. 95.4% confidence intervals are also shown.	241

8.9	Estimates for the durations of major archaeological intervals at Grotta Paglicci, extracted from the Partial (light grey) and Full (dark grey) Bayesian models (bottom half). Above, NGRIP $\delta^{18}\text{O}$ values (black; Rasmussen et al., 2014), % Arboreal pollen at Tenaghi Philippon (red; Müller et al., 2011), % Forest pollen at Ioannina (blue; Tzedakis et al., 2002), and % Forest pollen at Lago Grande de Monticchio (green; Allen et al., 1999) are plotted for comparison. Greenland Interstadials 8 to 2 are shown in grey and labelled, along with their uncertainties (Rasmussen et al., 2014).	242
8.10	Comparison of existing and new calibrated dates from stratigraphic layers at Grotta Paglicci.	244
8.11	New radiocarbon determinations made in this study, calibrated and modelled by Full and Partial Bayesian models. Except some inconsistencies between probability densities across Layer 23 and 21A1, the two models show excellent agreement.	247
8.12	Estimates for the emergence of the Epigravettian at Paglicci from Partial (light grey) and Full (dark grey) models. Corresponding 95.4% confidence intervals are shown below for both models. . . .	248

List of Tables

2.1	Predictor and response variables used in statistical downscaling. . .	28
2.2	Köppen-Geiger climate classification scheme.	30
2.3	Palaeoenvironmental records used for validating downscaling predictions over MIS 3. Column "Resolution" refers to the approximate chronological resolution per sample in years, as estimated by Fletcher et al. (2010).	31
2.4	Feature selection results for Temperature (T_{obs}) and Precipitation (P_{obs}) response variables. Column "N" gives the number of features for each model. The first row shows results for models built using all features, as selected by BIC. For rows below, the three models with the lowest RMSE for each number of features are shown. BIC and RMSE for each feature combination are shown, with σ_{RMSE} and SE calculated across spatial folds. RMSE for GAMs with features selected by OSE are shown in bold.	34
3.1	Results of alignment by DTW to downscaled temperature and precipitation predictions for the three pollen records considered. Column "Length" shows the length of the records between 60 and 30 kyr BP (MIS 3), and "Res" their resolution over the same period. L is the elasticity parameter used in the MVM step pattern. Columns "Dist" and "N Dist" refer to the DTW and normalised DTW distances in alignment, respectively.	78
4.1	Parameters and associated values used in this study.	101
4.2	Definitions of terms used in this study.	104

4.3	Example alignment problem for Pearson correlation and DTW, as shown in Figure 4.4, with one ring removed at the 5th (near-beginning), 15th (middle), and 25th (near-end) indices. Pearson correlation (r) and associated statistics (p , t -Test) are shown, along with DTW distance (C), normalised DTW distance (α), and the DTW adjusted t -Test score. Alignment indices (i.e., solutions) for both methods are also shown. The correct alignment is to index 51. Where the middle (15th) index is removed, Pearson correlation fails to find the correct solution, with a low r value and failed t -Test. DTW finds the correct solution (within one index) for all removed rings, passing the t -Test threshold.	123
5.1	Previous studies using ABOx-SC and ABA-SC pretreatment methods on paired samples. Column N relates to the number of samples dated by both ABOx-SC and ABA-SC.	142
5.2	Samples used in the experiment and their previous ABA and ABOx-SC radiocarbon determinations. ABA dates were made between 1996 and 2008 (column Year), while ABOx-SC dates were made 2007/8. χ^2 values for paired dates are given in the final column (df = 1, N = 2, $p < 0.05$)	148
5.3	New radiocarbon measurements produced in this experiment for ABA (2023) and ABA-SC pretreatment methods, for archaeological samples from Fumane and Maninjau background standards, along with stable carbon isotope ($\delta^{13}\text{C}$) and percent carbon (%C) measurements. ABOx-SC measurements made in 2007/8 are also shown for comparison.	157
5.4	The effect of precombustion on ABA-treated samples. Column headings: SC'd = amount of material precombusted (mg); SC Yield = amount of material remaining after precombustion (mg); % Change = percentage mass change; $\Delta\%C$ = change in percent carbon measurement after precombustion; $\Delta\delta^{13}\text{C}$ = change in stable carbon isotope ratio after precombustion; χ^2 = Chi-squared metric for ABA and ABA-SC dates (df = 1, $p < 0.05$); R = proportional effect of precombustion on ABA samples for contaminant removal, relative to the inferred amount removed by ABOx-SC.	159
5.5	Original and updated (2023) radiocarbon measurements for samples pretreated by ABA.	162
6.1	Archaeological samples used in experiment from Grotta di Fumane, previously dated by ABA and ABOx-SC methods. χ^2 statistics for paired dates are given in the final column (df = 1, N = 2, $p < 0.05$).	178

6.2	Chemical protocols for ABA, ABOx, AOX, and ABl pretreatment methods. Between all steps samples are washed at least three times with ultra pure water.	179
6.3	Radiocarbon determinations for samples used in experiment. ABl and ABl-SC dates were produced for this experiment, while ABA and ABOx-SC dates were previously produced at the ORAU.	182
8.1	Previous radiocarbon determinations for samples from Grotta Paglicci.	221
8.2	New radiocarbon determinations for samples from Grotta Paglicci. Calibrated age ranges are given at 95.4% confidence intervals for unmodelled dates, the Partial Bayesian model using only the new radiocarbon measurements made in this study, and the Full Model using all available radiocarbon dates from Paglicci.	235

List of Abbreviations

ABA-SC		Acid-Base-Acid with Stepped-Combustion
ABA		Acid-Base-Acid
ABOx-SC		Acid-Base-Oxidation with Stepped-Combustion
AIC		Akaike Information Criterion
AMOC		Atlantic Meridional Overturning Circulation
AMS		Accelerator Mass Spectrometry
AOx-SC		Acid-Oxidation with Stepped Combustion
BIC		Bayesian Information Criterion
calBP		Calibrated Years Before Present (CE 1950)
CE-SC		Cellulose Extraction with Stepped-Combustion
CI		Campanian Ignimbrite
CRU		Climatic Research Unit, University of East Anglia
DEM		Digital Elevation Model
DO		Dansgaard-Oeschger
DTW		Dynamic Time Warping
EPS		Expressed Population Signal
FTIR		Fourier Transform Infrared Spectroscopy
GAM		Generalised Additive Model
GCM		General Circulation Model
GI		Greenland Interstadial
GLK		Gleichläufigkeitskoeffizient
GS		Greenland Stadial
HadCM3B	. .	Hadley Centre Coupled Model 3 Bristol
hypy		Hydrogen Pyrolysis
IRMS		Isotope Ratio Mass Spectrometer

kyr BP		Thousands of Years Before Present (CE 1950)
MIS		Marine Isotope Stage
MVM		Minimal Variance Matching
NGRIP		North Greenland Ice Core Project
ORAU		Oxford Radiocarbon Accelerator Unit
OREC		Oxidation Resistant Elemental Carbon
OSE		One Standard Error Rule
OSL		Optically Stimulated Luminescence
QQ		Quantile-Quantile
RMSE		Root Mean Squared Error
RWI		Ring Width Indices
SRTM-30 Plus		Shuttle Radar Topography Mission

1

Introduction

Palaeoenvironmental variability is frequently employed to understand and contextualise the human past. Dramatic fluctuations in the earth's climate system have unfolded over the past ~ 125 thousand years (kyr), bringing with them significant changes in the earth's surface temperatures, precipitation patterns, ice sheet volumes, and sea level. Over the same period, the distribution of our species expanded from the African continent, into the high latitudes of Eurasia, across large bodies of water to Sahul, and over the now submerged steppic plain of Beringia into the varied landscapes of the Americas. The tool kit and material expressions of *Homo sapiens* populations underwent significant and diverse changes, population size increased, and the capacity of our species to survive in a vast range of ecological settings was confirmed. Within the same period, multiple species of our own genus disappeared; by 30 kyr ago, *sapiens* were the only surviving hominin species on earth. These changes in human behaviour, technology, and distribution unfolded over millennia, against a backdrop of substantial changes to the earth's climate system. Despite decades of research, it is unclear whether millennial-scale climate change was a major driver of any of these uniquely human processes.

These questions have drawn particular interest in the study of Palaeolithic Europe, where a long history of excavation has produced a detailed archaeological

record, that can be compared to many high-resolution records of past climate variability from the same region. The European Palaeolithic presents a wealth of opportunities to explore, forward, and test theories for causal relationships between archaeological change and environmental processes. The arrival of *Homo sapiens* and concomitant disappearance of Neanderthal populations in Europe is frequently explained and understood with reference to climate variability, around themes such as adaptation or eco-cultural niches. Technological and cultural shifts of later periods of the Upper Palaeolithic, such as the Aurignacian and Gravettian, have similarly been forwarded as behavioural responses to changing environmental conditions and “climatic stress.”

This broad field is truly interdisciplinary: archaeology reveals the material evidence of human presence, behaviours, and cultural change; the social sciences frame our understanding of these symbolic expressions, materialities, and technologies; environmental and climate sciences provide evidence for changes in the earth’s climate system and their footprint; and biology and zoology inform the study of population dynamics, movement, and mixing, as well as the impact human populations had on other species. The science of chronology, meanwhile, is the fundamental framework by which these complex fields are brought together. Crucially underwriting the entire field, chronology enables the study of the relationship between human and environmental processes, and their evolution over time.

Drawing parallels between records for the Late Pleistocene, be they archaeological, palaeoenvironmental, or otherwise, demands confidence in their chronologies. However, significant challenges surround the precision, accuracy, and resolution of Late Pleistocene and Palaeolithic records and dates. The sparsity of chronological information can preclude the alignment of changes in the archaeological record to environmental phenomena, which in turn creates a barrier to addressing direct causality between such processes.

This thesis explores these challenges, focusing on the themes of *Climate* and *Chronology*. The common aims of the chapters are to illuminate and alleviate current problems relating one or both of these themes. The research is centred

on archaeological change in Europe and its environmental backdrop during the Late Pleistocene, and employs methods such as Machine Learning statistics, signal processing algorithms, and radiocarbon measurements to understand the relationship between these processes in more detail. The research contributes to the fields of palaeoenvironmental and archaeological science by novel applications of numerical methods to better understand existing records, as well as the production of new data.

Background

The arrival *Homo sapiens* in Europe and emergence of the technologies of the Upper Palaeolithic occurred during a period of pronounced climatic variability known as Marine Isotope Stage 3 (MIS 3; ~ 60 to 30 thousand years before present; kyr BP), involving dramatic and abrupt changes to the earth climate system. Deciphering the impact of these changes on Palaeolithic populations in Europe requires a clear picture of *how* they unfolded over the northern hemisphere, along with accurate and precise estimates for *when* archaeological and palaeoenvironmental changes occurred.

At continental scales, current evidence supports *Homo sapiens*' arrival in Europe between 50 and 40 thousand calibrated radiocarbon years before present (ky calBP; [Higham et al., 2014](#)). The vast majority of the dates that contribute to this estimate are radiocarbon measurements. These chronologies suggest the process of *Homo sapiens* dispersal in Europe unfolded over millennia, along with gradual but fundamental technological changes to the *sapiens* tool kit: Initial Upper Palaeolithic industries and *Homo sapiens* remains at Bacho Kiro in Bulgaria are dated to c. 46.9 ky calBP ([Fewlass et al., 2020](#)); Uluzzian industries emerged in Greece and Italy by c. 45 ky calBP ([Douka et al., 2014](#)); and Aurignacian industries were present in Iberia by c. 44 ky calBP ([Marín-Arroyo et al., 2018](#); [R. Wood et al., 2014](#)). Taken broadly, a chronological overlap of several thousand years between these earliest arrivals of *Homo sapiens* and the disappearance of Neanderthal populations by around 40 ky calBP can be assumed. Evidence for

interbreeding between Neanderthal and *Homo sapiens* populations confirms that, for some periods of MIS 3, these populations were present in the same region (Fu et al., 2015; Slon et al., 2018).

Many studies have paid particular focus to abrupt warming episodes and temperature oscillations during MIS 3 to explain these processes (Andel, 2005; Badino et al., 2020; Müller et al., 2011; Staubwasser et al., 2018; Zilhão et al., 2017). The oscillations, known as Dansgaard-Oeschger (DO) cycles, are evidenced by fluctuating stable oxygen isotope ratios in Greenlandic ice cores, from which atmospheric temperatures over Greenland can be confidently inferred (Kindler et al., 2014). DO cycles are characterised by abrupt warming episodes from stadial conditions, followed by gradual cooling through interstadial periods of several millennia, before an abrupt return to stadial conditions (Dansgaard et al., 1993; Rasmussen et al., 2014).

Many European palaeoenvironmental records appear to reflect similar oscillations to the DO cycles observed in Greenland (Budsky et al., 2019; Müller et al., 2011; Sánchez Goñi et al., 2018; Shackleton et al., 2000; Staubwasser et al., 2018), but their timings and spatial footprints are not clear. Some European speleothem records show abrupt changes in their stable isotope ratios that were synchronous with Greenland, but others suggest relative climatic stability over MIS 3 (Adolphi & Muscheler, 2016; Columbu et al., 2020; Corrick et al., 2020). Pollen records show expansions and contractions of woodland and steppic species for MIS 3, again suggesting parallel temperature oscillations unfolded in Europe (Allen et al., 2000; Tzedakis et al., 2002), but cold and arid signals are more pronounced in some records than others (Badino et al., 2020; Fletcher et al., 2010; Müller et al., 2011; Sánchez Goñi, 2022). Based on their independent chronologies, some pollen signals appear to have been out of phase with the Greenland record, as well as each other (Blaauw et al., 2010). Building robust radiocarbon chronologies for pollen records has been particularly difficult, due to reservoir effects, sediment mixing, or a lack of material for dating altogether. Synthesising these records to a coherent

picture of palaeoenvironmental change in Europe over MIS 3 is a challenging task, involving considerable uncertainties.

European Palaeolithic archaeology has suffered from similar chronological challenges, particularly for sequences close to the radiocarbon boundary (around 55 thousand years before present). Although many radiocarbon measurements exist for the archaeology of MIS 3 and 2 (e.g., the Radiocarbon Palaeolithic Database; [Vermeersch, 2020](#)), reliable and coherent chronological frameworks for the period have been difficult to produce, at both site and regional levels. Over the past two decades it has become increasingly apparent that many previous radiocarbon determinations significantly underestimated the true age of Upper Palaeolithic samples, sites, and shifts in the archaeological record ([Douka et al., 2014](#); [Higham, Brock, et al., 2009](#); [Higham, 2011](#); [R. Wood et al., 2014](#)). Very small amounts of exogenous or modern ^{14}C contamination can result in dating inaccuracies of thousands of years ([R. Wood, 2015](#)). Typically, the pattern has been that the chronologies are “pushed back” in time by their revisions, with improved contaminant removal prior to radiocarbon measurement. Modern instruments have improved measurement precision, meanwhile, and updates to calibration curves, in combination with Bayesian methods, can further constrain chronologies. These developments provide clear opportunities for updating chronologies of the Palaeolithic, but many “old” radiocarbon measurements continue to contribute to archaeological discussion, modelling, analysis, and interpretation, without assessment or critique.

Given the challenges surrounding both archaeological and palaeoenvironmental chronologies, it is sometimes difficult to make links between them with any confidence. Figure 1.1 demonstrates this problem. An example radiocarbon date with a typical error for the Palaeolithic period (40000 ± 500 ^{14}C years) is shown, along with temperature reconstructions over Greenland by [Kindler et al. \(2014\)](#). Many of the abrupt warming or cooling episodes observed over Greenland occurred over periods shorter than the ranges of most radiocarbon dates for this period, while the chronological uncertainty of the Greenland record adds further complexity to this problem ([Rasmussen et al., 2014](#)). A single date for the Palaeolithic period

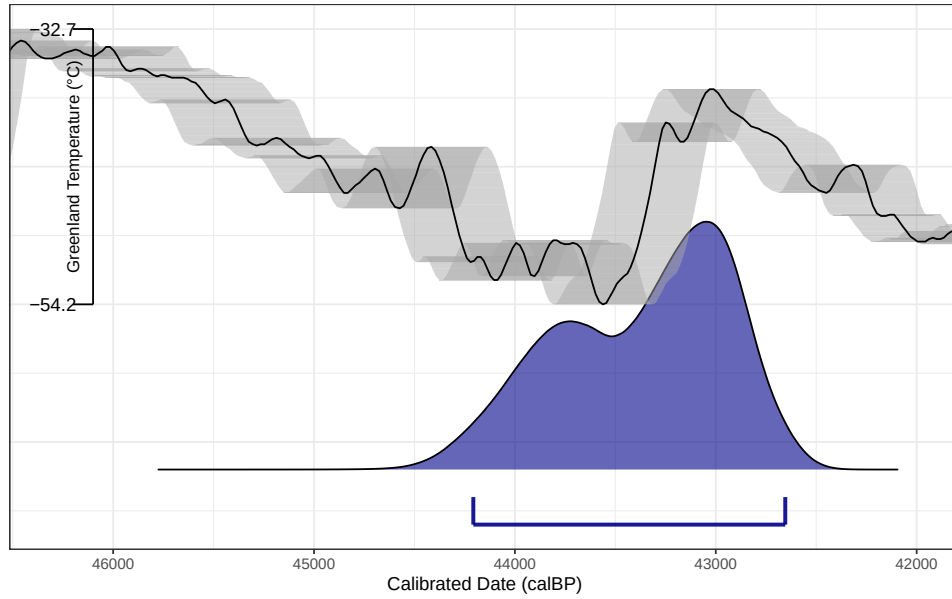


Figure 1.1: An example calibrated radiocarbon measurement of 40000 ± 500 ^{14}C years (blue shading), calibrating to 44207-42654 calBP at the 95.4% confidence interval (blue line). The calibrated date is plotted alongside the Greenlandic temperature reconstruction (black line, Kindler et al. 2014) and its associated chronological uncertainty (grey shading; Rasmussen et al., 2014). Considering the uncertainties of the radiocarbon date and Greenland record, it is not clear whether the date relates to a warm, cold, or transitional period of MIS 3.

might cover cold, warm, and transitional periods; although the Greenland record is frequently employed to explain archaeological change over the Palaeolithic, it is often unclear whether these changes relate to periods before, during, or after stadial, interstadial, or transitional conditions over MIS 3.

Whilst it has often been suggested that variability during MIS 3 constrained or enabled periods of hominin dispersal and population contraction (Müller et al., 2011; Staubwasser et al., 2018; Stewart et al., 2019; Timmermann, 2020), or drove technological innovation and cultural change (Conard & Bolus, 2003; Teller & Conard, 2022), any causal relationship between these processes is very difficult to infer with confidence. This is in part due to the sparsity, uncertainty, and resolution of both archaeological and palaeoenvironmental records from this period. Many relevant records lack the precision and accuracy in their chronologies to allow direct and meaningful parallels to be drawn between them. A still broader issue is the environmentally deterministic assumption that various archaeological shifts

during the Palaeolithic might be explained by environmental change at all.

Research Aim and Objectives

The aim and objectives of this research are centred around the themes of *Chronology* and *Climate* during MIS 3. Palaeolithic archaeology is frequently contextualised by environmental variability, but the chronological constraints and sampling resolutions of both palaeoenvironmental and archaeological records are often too limited to allow links to be drawn between them confidently. The aim of this thesis is to improve chronological estimates for both archaeological and palaeoenvironmental changes during the Late Pleistocene, using: numerical methods and simulations, such as General Circulation Models, Bayesian statistics, and signal processing algorithms; and laboratory-based research, producing new radiocarbon dates and exploring experimental sample pretreatment methods. Under this aim, the objectives of this thesis are:

Objective 1: Develop robust statistical methods to downscale General Circulation Models and reconstruct the spatial impact of Dansgaard-Oeschger events across Europe during MIS 3

Objective 2: Evaluate modelled temperature and precipitation variability by downscaled General Circulation Models by comparing their predictions to palaeoenvironmental records

Objective 3: Adapt and apply signal processing algorithms to assimilate modelled and observed data and produce new chronologies of palaeoenvironmental change over MIS 3

Objective 4: Explore and design new methods for radiocarbon pretreatment

Objective 5: Produce new radiocarbon dates by the most robust pretreatment methods to provide updated chronologies of Palaeolithic cultural change in Europe, further constrained by Bayesian statistical modelling

Objective 6: Compare updated estimates for the timings of archaeological change to palaeoenvironmental records and assess their relationship

Thesis Structure and Chapter Outline

All chapters in this thesis are presented as independent studies, with the exceptions of: Chapter 3, where predictions made in Chapter 2 are used in combination with existing data and a signal processing algorithm; and Chapter 7, where the combined results and implications of radiocarbon experiments (Chapter 5 and Chapter 6) are discussed. Except Chapter 7, all chapters feature an introduction with an assessment of relevant literature, a description of methods, a presentation of results, and a discussion of the study's implications.

Chapter 2: *Statistical Downscaling of a GCM Simulation Over Europe During MIS 3* presents a high-resolution simulation of climate variability over Europe between 60 and 30 thousand years before present (kyr BP). A General Circulation Model (GCM) simulation for MIS 3 provides the foundation for this study (Armstrong et al., 2019), with its spatial resolution improved by a method known as statistical downscaling. Machine Learning methods are employed to build, tune, and cross-validate statistical models of climate variability, before applying the models to simulations to produce high-resolution temperature and precipitation reconstructions over Europe during MIS 3. Downscaled predictions are compared to relevant palaeoenvironmental proxies (Figure 1.3), with results highlighting the difficulty of reconciling models and observations in palaeoclimate research.

Building on the results from Chapter 2, Chapter 3: *Synchronised Proxy Records for MIS 3* applies a signal processing algorithm known as Dynamic Time Warping (DTW) to three pollen records of environmental variability, to synchronise their signals with downscaled GCM predictions at their respective locations in Europe (Figure 1.3). The resulting chronologies for each record are compared to their existing age-depth models, highlighting discrepancies between existing age models and the realigned signals.

Another palaeoenvironmental application of DTW is considered in Chapter 4: *Dendrochronological Applications of Dynamic Time Warping*. The problem of cross-dating in dendrochronology involves aligning a sample to the best-match

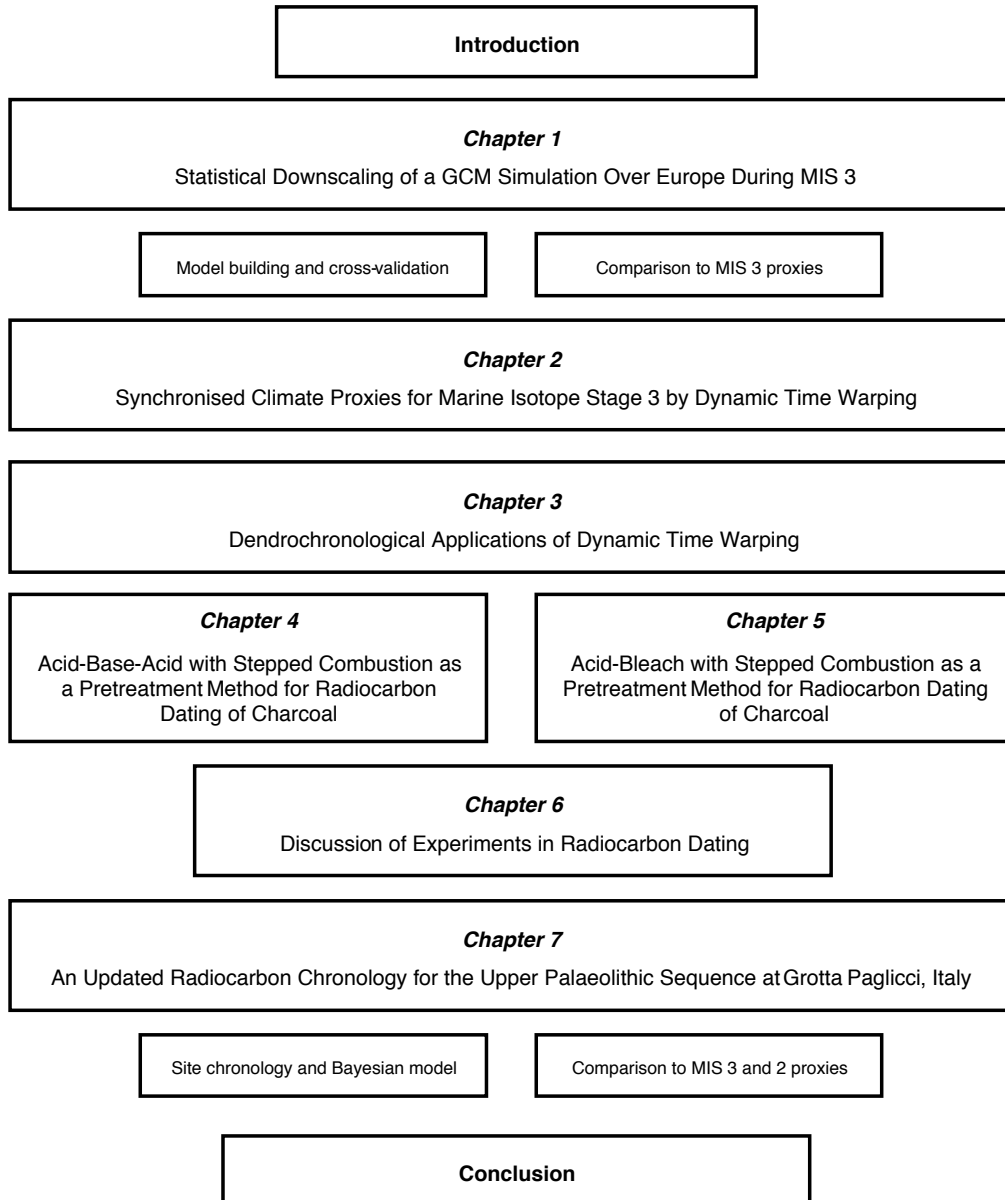


Figure 1.2: Chapter structure.

indices of a master sequence, an averaged record of ring widths for a given region and tree species. An example tree ring data set from Towai Farm, New Zealand, is used to create thousands of samples of varying length (Palmer et al., 2016), with rings added and removed to simulate samples that are problematic for cross-dating by traditional methods. Chapter 4’s experiment shows that DTW performs exceptionally well in this task, being more resilient to perturbations in the sample than the conventional method of alignment by Pearson correlation.

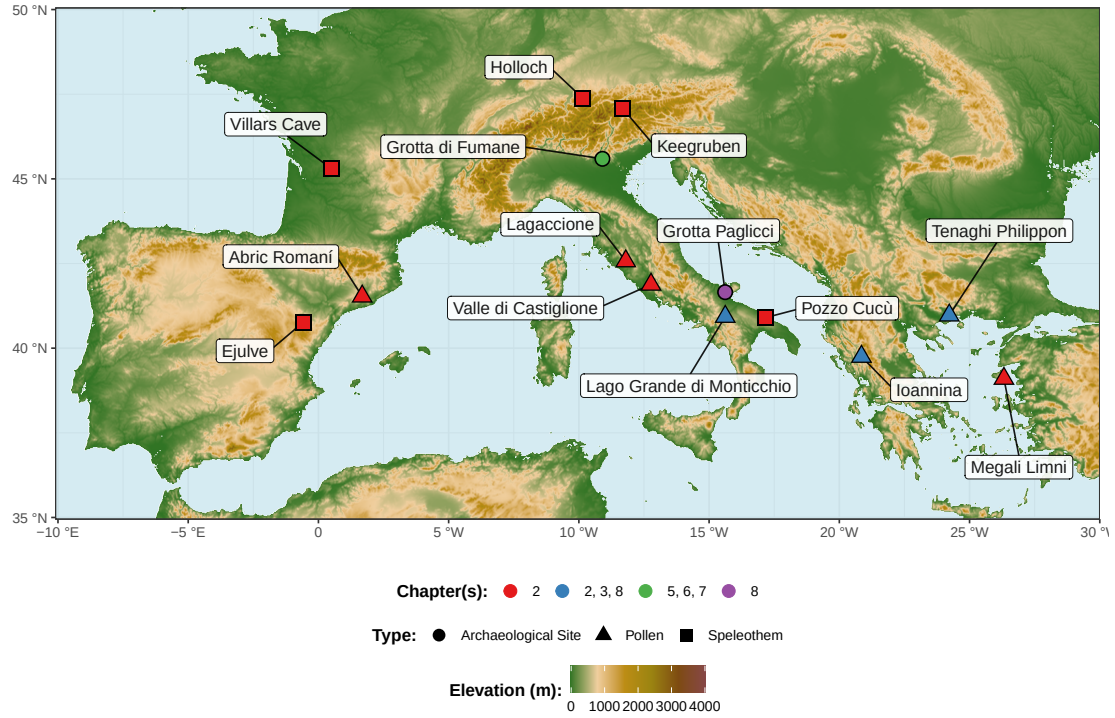


Figure 1.3: Locations of palaeoenvironmental records and archaeological sites considered in this thesis.

Chapter 5 and Chapter 6 present laboratory-based experiments focused on radiocarbon dating of charcoal. Two experimental pretreatment methods are designed, both intended to remove contamination from charcoal prior to radiocarbon measurement, using charcoal samples from a key Upper Palaeolithic site in Italy, Grotta di Fumane (Figure 1.3). Chapter 5: *Acid-Base-Acid with Stepped Combustion (ABA-SC) as a Pretreatment Method for Radiocarbon Dating of Charcoal* adapts an existing pretreatment method (ABA) by adding a stepped combustion step to the protocol. Chapter 6: *Acid-Bleach with Stepped Combustion (ABL-SC) as a Pretreatment Method for the Radiocarbon Dating of Charcoal* explores an alternative oxidation reagent in aggressive charcoal chemical pretreatment. The implications of the combined results of these experiments are then discussed in Chapter 7: *Discussion of Experiments in Radiocarbon*.

Finally, a case study exploring the relationship between archaeological changes during the Upper Palaeolithic and large-scale environmental changes over MIS 3 and 2 is presented in Chapter 8: *An Updated Radiocarbon Chronology for the*

Upper Palaeolithic Sequence at Grotta Paglicci, Italy. This chapter presents 26 new radiocarbon dates on charcoal from a key Upper Palaeolithic site for understanding the Aurignacian and Gravettian industries on the Italian Peninsula, Grotta Paglicci (Figure 1.3). Bayesian models provide estimates for the timings of cultural transitions the site, which are then compared to records of palaeoenvironmental variability. Results show that both the Aurignacian and Gravettian persist at Grotta Paglicci through periods of pronounced variability. As found for previous chapters, the resolution and chronological constraints of relevant records are too poor to provide deep insight into causal relationships between the emergence of certain behaviours or technologies and large-scale environmental change.

2

Statistical Downscaling of a GCM Simulation Over Europe During MIS 3

Abstract

Repeated, dramatic, and abrupt climate oscillations are recorded in Greenland ice core records during Marine Isotope Stage 3 (MIS 3; ~ 60 to 30 kyr BP). European palaeoenvironmental records suggest that the ecological signals of these oscillations were spatially complex, and may not have been synchronous with those over Greenland. The effect of this environmental variability on hominin populations and their distributions across Europe, if any, are currently unclear.

General Circulation Models (GCMs) can provide alternative numerical reconstructions of palaeoenvironmental change, with the distinct advantage of complete spatial coverage, albeit at coarse resolution. To increase their spatial resolution, GCM simulations can be “downscaled” using statistical methods. Although downscaling circumvents the problem of low spatial resolution, it is essential that downscaled predictions and associated statistical methods are evaluated appropriately. Models should be built and cross-validated robustly, with predictions compared to existing palaeoenvironmental records.

In this chapter, a continuous GCM time series for MIS 3 is used to downscale temperature and precipitation over Europe by Generalised Additive Models (GAMs; [Armstrong et al., 2019](#)). Models are constructed with a k -fold spatial cross-validation scheme for feature selection, before testing their accuracy on a hold-out dataset. High predictive accuracy for temperature is identified through this scheme, but comparatively poor performance is apparent for precipitation. GAMs are used to downscale GCM surfaces over MIS 3, at 0.16° spatial resolution and 100-year time steps. The resulting palaeoclimate reconstruction is validated by comparison to terrestrial palaeoenvironmental records.

Downscaled time series for MIS 3 closely resemble the Greenland ice core record, due to the correction to the GCM simulations made by [Armstrong et al. \(2019\)](#). However, downscaled predictions and proxy observations of palaeoenvironmental change over Europe show poor agreement, particularly where resolutions of the latter are low. This raises questions over the application of GCM-derived predictions to archaeological questions, such as their use in species distribution or eco-cultural modelling. Overall, the validation scheme mirrors previous comparisons between these palaeoenvironmental signals and the Greenland records, highlighting a number of challenges relating to the chronological calibration, alignment, and tuning of palaeoenvironmental proxies for MIS 3. These are considered and discussed further in [Chapter 3](#).

2.1 Introduction

The archaeology of the Middle and Upper Palaeolithic in Europe is frequently contextualised by palaeoenvironmental variability. Stable oxygen isotope ratios ($\delta^{18}\text{O}$) measured through Greenlandic ice core records are often compared to Middle and Upper Palaeolithic assemblages, sites, and sequences, given their high sampling resolution and confident association with changes in atmospheric temperatures (Kindler et al., 2014; Rasmussen et al., 2014). Variability in Greenlandic ice core records likely reflects large-scale (i.e., hemispheric) shifts in the earth’s climate system, and these shifts are particularly pronounced during Marine Isotope Stage 3 (MIS 3; 60 to 30 kyr BP) (Malmierca-Vallet & Sime, 2023).

Repeated episodes of abrupt warming to interstadial conditions evidenced by Greenland records are known as Dansgaard-Oeschger (DO) events (Dansgaard et al., 1993). DO events are thought to be indicative of large-scale northern hemispheric temperature variation, closely linked to changes in the strength of the Atlantic Meridional Overturning Circulation (AMOC) (Broecker & Denton, 1990; Henry et al., 2016; Izumi et al., 2023; Menviel et al., 2020; Voelker, 2002), characterised by a sharp temperature increase, followed by an interstadial period of relative stability characterised by warm temperatures with gradual cooling, before abrupt cooling to stadial conditions (Izumi et al., 2023; Kindler et al., 2014). DO events have been intensely studied and modelled since their first identification (Dansgaard et al., 1993), and their associated cycles have been identified in multiple ice core proxies and analyses. These include $\delta^{18}\text{O}$ measurements (Dansgaard et al., 1993), dust content (Ruth et al., 2003), and greenhouse gas concentrations (Schilt et al., 2010). Together, these proxies suggest that DO events were large-scale atmospheric processes encompassing multiple components of the earth system. However, although many palaeoenvironmental records from the northern hemisphere also show DO-like patterns in their signals, the spatial footprint of these events has proven difficult to reconstruct (Blaauw, 2012; Fletcher et al., 2010). Sparse and inconsistent sampling resolutions for many European records, coupled with poor

chronological constraints, have prevented the building of a coherent picture of DO events and their ecological impact.

General Circulation Models (GCMs) have been the foundation of many studies on DO events (Boers, 2018; Izumi et al., 2023; Menviel et al., 2020). GCMs are numerical representations of the climate system, used to model its evolution over time: given a set of initial conditions, such as greenhouse gas concentrations or orbital parameters, a GCM provides a simulation of the earth system and its response to these perturbations. GCMs are well-known for projecting to future scenarios of climate change (Intergovernmental Panel on Climate Change, 2014, 2023), but they also have valuable application in palaeoenvironmental and archaeological sciences. As well as illuminating the mechanisms of DO events, GCMs offer an alternative method of palaeoenvironmental reconstruction, providing an opportunity to circumvent some of the challenges associated with European terrestrial observations.

Palaeoclimate modelling by GCMs is an active field of research in and of itself, where simulations of past states of the earth’s climate have improved understanding of the earth system and its processes (Haywood et al., 2019; Valdes et al., 2017). Quantitative estimates for feedback mechanisms, climate sensitivity, or interactions between spheres of the climate system can be explored and tested through palaeoclimate modelling. Palaeoclimate modelling uniquely allows for model-data comparisons against palaeoenvironmental observations, such as the Greenland record, and the opportunity to test hypotheses in climate dynamics (Braconnot et al., 2012). Given the wealth of data provided by palaeoclimate simulations, their output has been adopted in other fields, such as archaeology, palaeoecology, and palaeobiogeography (Banks et al., 2013; Haywood et al., 2019; Pilowsky et al., 2022; Varela et al., 2011; Willerslev et al., 2014). Over the past two decades, GCMs have been regularly applied to questions in Palaeolithic archaeology, particularly to discuss hominin “niche” dynamics and human dispersals, for example (Andel, 2005; Banks et al., 2006, 2013; Banks, 2017; Bradshaw et al., 2019; Eriksson et al., 2012; Franklin et al., 2015; Giampoudakis et al., 2017; Larrasoana et al., 2013; Timmermann, 2020; Timmermann & Friedrich, 2016).

A significant limitation of using GCM simulations for palaeoenvironmental reconstruction is that their spatial and chronological resolutions are often limited. Palaeoclimate simulations are typically between 1° and 10° of latitude and longitude per terrestrial grid cell, being too coarse for most archaeological applications except perhaps at global scales (e.g., [Eriksson et al., 2012](#)). To alleviate this constraint, an intermediate transformation can be applied to the GCM to “zoom in” and increase the spatial resolution of its output, through a process known as “downscaling” ([Beyer et al., 2020](#); [Maraun & Widmann, 2018](#); [Vrac et al., 2007](#)). Downscaling involves quantifying the relationship between coarse-resolution GCM surfaces and modern climate observations. This relationship is then applied to future or past simulations to increase their spatial resolution. If the resolution of the observations is high, so too are the downscaled data. Downscaling can be achieved by a number of methods, from simple “delta” corrections ([Armstrong et al., 2019](#); [Beyer et al., 2020](#)), to more complex statistical methods ([Sachindra et al., 2018](#)).

Although the spatial resolution of a GCM can be improved by downscaling, the problem of chronological resolution is more challenging. It is rare that a GCM is run to produce continuous simulations for a period as long as MIS 3, for example. In archaeology, previous studies have used “cold” and “warm” simulations as proxies for the extremes of climatic oscillations over the Late Pleistocene, with species distributions and changing niche dynamics then inferred by comparisons between these simulations ([Andel, 2005](#); [Banks et al., 2013](#)). For MIS 3, this may fail to address the variance and ecological complexity of distinct and separate DO episodes, as implied by considerable variability in the signals in palaeoecological records. Furthermore, the rate of climatic change is overlooked in such an approach, which may be pertinent to questions of hominin adaptation.

There are also “downstream” solutions to limited chronological resolution in palaeoclimate modelling. The HadCM3B GCM has provided 62 palaeoclimate simulations at intervals over the Late Pleistocene ([Singarayer & Valdes, 2010](#); [Valdes et al., 2017](#)), and these simulations have been used in a variety of palaeoecological and archaeological applications ([Beyer et al., 2021](#); [Eriksson et al., 2012](#); [Jennings](#)

et al., 2015; Krapp et al., 2021; Saltr  et al., 2016). Armstrong et al. (2019) designed an innovative method to combine these “snapshot” simulations, to produce a continuous time series of GCM output for the period between 60 and 0 kyr BP, at monthly resolution. DO variability at millennial scales was superimposed on the time series, using a correction surface scaled by the Greenland record (Kindler et al., 2014). The resulting time series was then downscaled by the Delta Method to 0.5  resolution, providing a continuous reconstruction of climate variability over the northern hemisphere that closely resembles the Greenland record.

In validating the model, Armstrong et al. (2019) found good agreement with 21 temperature reconstructions from ice cores across the northern hemisphere. However, other than by records from Greenland, validation was limited to the period between 24 and 0 kyr BP. 18 of the 21 ice core reconstructions did not cover MIS 3, and the remaining three were all Greenlandic, which, unsurprisingly, correlate well with the modelled data. Only two lower-latitude and terrestrial temperature reconstructions were used in validation, one from Central China and the other from Alaska. Both showed poor agreement to the HadCM3B-based estimates, particularly over MIS 3. Armstrong et al. (2019) note that a more advanced (i.e., statistical) downscaling technique might improve model-data agreement, but arguably, there was also insufficient validation for adequate evaluation.

Building on the study by Armstrong et al. (2019), this chapter aims to statistically downscale and further validate the HadCM3B simulations, specifically for MIS 3. In this study, the time series of interpolated and DO-corrected GCM surfaces produced by Armstrong et al. (2019) is downscaled by Machine Learning methods to produce high-resolution reconstructions of palaeoclimatic changes over Europe. Five relevant GCM variables are considered for feature selection, along with four topographic variables. Generalised Additive Models are used for downscaling, applied by a k -fold spatial cross-validation scheme and automated feature selection process. Temperature and precipitation variables are downscaled to 0.16  resolution over the study region, at 100-year time steps between 60 and 30 kyr BP. Predictions are compared to relevant European palaeoenvironmental records for validation.

Finally, the key uncertainties and directions for future research in this field are explored and highlighted.

2.2 Materials and methods

2.2.1 Study Region

The study region was defined as the area between longitudes -10° E and 27° W, and latitudes 35° and 50° N (Figure 2.1). This area was chosen to include: the Iberian and Italian peninsulas, both key regions for understanding the Middle to Upper Palaeolithic Transition in Europe due to their rich archaeological records (Haws et al., 2020, 2021; Higham et al., 2011; Zilhão et al., 2006); the Early Upper Palaeolithic site of Bacho Kiro to the east of the study region, currently showing the earliest evidence for *Homo sapiens* in Europe (Fewlass et al., 2020); a wide range of biomes and ecological systems, such as mountainous, arid, forested, and temperate zones, in order to encompass a large range of temperature and precipitation values in model training (see Section 2.2.3.2); and a sufficient number of terrestrial palaeoenvironmental records, at sufficient chronological resolution, to compare downscaled results to (see Section 2.2.5). All geospatial operations are performed in the native spatial projection of the response variables, WGS84.

2.2.2 General Circulation Model

This study builds on a time series of palaeoclimate simulations between 60 and 0 ky BP produced by the Hadley Centre Coupled Model 3 Bristol (HadCM3B; Armstrong et al., 2019; Valdes et al., 2017). HadCM3B is a coupled GCM with 3D dynamical atmosphere and ocean components and an atmospheric spatial resolution of 3.75° longitude x 2.5° latitude x 19 vertical levels (Valdes et al., 2017). The time series produced by Armstrong et al. (2019) is based on 42 HadCM3B “snapshot” simulations of the earth climate system during the Late Pleistocene, run at 2 ky intervals between 60 and 22 kyr BP and 1 ky intervals between 21 and 0 kyr BP. Variable initial conditions driving each simulation were greenhouse gas concentrations, orbital parameters, ice sheet extent, and sea-level, derived from a combination of observations and modelling methods. The snapshots were spline interpolated to 1-year resolution by Armstrong et al. (2019), with millennial scale

variability reflecting DO cycles superimposed onto the yearly time series using a correction surface and scaling factor from a separate freshwater forcing simulation and the NGRIP ice core record (Kindler et al., 2014).

The only differences between the GCM data used in this study and the data published by Armstrong et al. (2019) is that the latter includes added interannual variability (i.e., GCM stochasticity) and applies a “delta method” downscaling (described below) to increase the data’s spatial resolution. The GCM surfaces used as predictors for MIS 3 in this chapter were spline interpolated and DO-corrected without any further transformation.

2.2.3 Statistical Downscaling Scheme

A number of downscaling methods have been established and applied to palaeoenvironmental contexts, ranging from relatively simple linear corrections, to more computationally intensive approaches using nested simulations, or “regional” climate models (Beyer et al., 2020; Engelbrecht et al., 2019). Generally, the latter is not feasible for generating continuous multi-millennial time series at high spatial and chronological resolution. “Delta” methods belong to the former category, representing a much simpler approach to the problem. Here, a brief overview of the delta method is given, before describing the statistical downscaling approach taken in this study.

Using the delta method, a high-resolution correction (Δ) surface for a given climate variable is calculated, then applied to past or future GCM surfaces to improve their resolution. For temperature, a grid cell’s Δ value is given by the difference between modern observations (T_{obs}) and smoothed (2-d interpolated) GCM simulation for the modern period (T_{g_0}) at that location. The delta correction can then be applied to other GCM time steps:

$$\Delta_T = T_{g_0} - T_{obs} \quad (2.1)$$

$$T_t = T_g(t) + \Delta_T \quad (2.2)$$

As precipitation (P) is bounded by zero, the ratio of modern observations (P_{obs}) to the GCM simulation for the modern period is used, rather than the difference, with the correction applied as a multiplication:

$$\Delta_P = \frac{P_{obs}}{P_{g0}} \quad (2.3)$$

$$P_t = P_g(t)\Delta_P \quad (2.4)$$

An assumption is that the delta corrections derived in this process have been constant through time, or at least are directly applicable to the past or future period of interest. Whilst the delta method has recently been argued to perform as effectively as more complex downscaling approaches (Beyer et al., 2020), the available validation methods for assessing this claim are reliant on robust chronologies and palaeoenvironmental records, which arguably become increasingly lacking for earlier periods of the Pleistocene. Another drawback of the delta method is that Δ can only be defined at locations where modern observations have been made. This means that corrections cannot be applied to areas now submerged due to sea level change, such as areas of the Adriatic Sea or North Sea.

Statistical downscaling using transfer functions is another relatively simple yet potentially powerful method of downscaling GCM output. Similar to the delta method, this approach involves identifying and quantifying relationships between large-scale atmospheric processes that are ideally well-represented by the GCM, and local patterns in climate variability. High-resolution observations of a given climate variable are treated as response variables in a statistical model, with GCM output and topographic information used as predictors. If statistical relationships between these variables can be confidently identified, the modelled functions can then be applied to the same GCM’s simulations of past or future climate, and high-resolution “downscaled” predictions for the respective period produced.

As for the delta method, an implicit and limiting assumption of the statistical downscaling approach is that any statistical relationship derived between predictor and response variables has been constant through time. Another broader limitation

of geostatistical models is their susceptibility to spatial autocorrelation (Meyer et al., 2019). However, an overriding advantage of this approach is that uncertainties and biases, as well as the overall performance of a model, can be more readily quantified and addressed than under the delta method, given the method’s foundation in statistical modelling and analysis (Katz, 2002).

To assess potential spatial biases and autocorrelative effects in models built for this study, models were cross-validated and tested on a “hold-out” data set (i.e., unseen data from within the model’s spatial domain). Validation methods are described in Section 2.2.3.4, and were applied in addition to validation by comparison to palaeoenvironmental records as described in Section 2.2.5.

2.2.3.1 Generalised Additive Models

Generalised Additive Models (GAMs) are one of the most frequently used models in statistical downscaling, due to their relative simplicity and capacity to capture non-linearity in relationships between response and predictor variables (Latombe et al., 2018; Levavasseur et al., 2011; Vrac et al., 2007; Woillez et al., 2014). In a GAM, the expected value of a response variable Y , given a set of J predictor variables $(X_1, \dots, X_J)$, is expressed as the sum of J non-parametric functions:

$$E(Y|X_1, \dots, X_J) = \sum_{j=1}^J f_j(X_j) \quad (2.5)$$

Each predictor’s function f_j is approximated by a third-order piece-wise polynomial, or cubic spline function, fitted to capture the relationship between X_j and Y . The “flexibility” of each spline is determined by the number of knots (k), which in turn determines the number of pieces in each polynomial.

The approach applied here followed Vrac et al. (2007) in GAM training, using $k = 12$ for all predictors. GAMs were trained for temperature and precipitation using data for all months (i.e., annual models were built, rather than separate models for each month). This avoids the potential for unrealistic jumps in seasonality from month-to-month in predictions, given changes in seasonality occur over longer scales

over the year (Vrac et al., 2007). Models were built using `mgcv`’s `bam` function in R and applied in the `caret` Machine Learning toolbox (M. Kuhn, 2016; R Core Team, 2020; S. N. Wood et al., 2015).¹

2.2.3.2 Response Variables

Present-day temperature and precipitation data used for model training were obtained from the Climatic Research Unit at the University of East Anglia (CRU; New et al., 2002). This global dataset was produced by interpolating observations made at irregularly spaced stations to 0.16° spatial resolution, as a function of latitude, longitude, and elevation. The resulting dataset represents mean temperatures and total precipitation for months averaged for the period 1961-1990, over all land areas excluding Antarctica. CRU temperature and precipitation observations, cropped to the study area, were treated as response variables in all statistical models described below ($N = 134,400$; 12 observations per grid cell). For temperature, a Gaussian distribution was assumed in model training, whilst a Gamma distribution with log-link was used for precipitation (Figure 2.1).

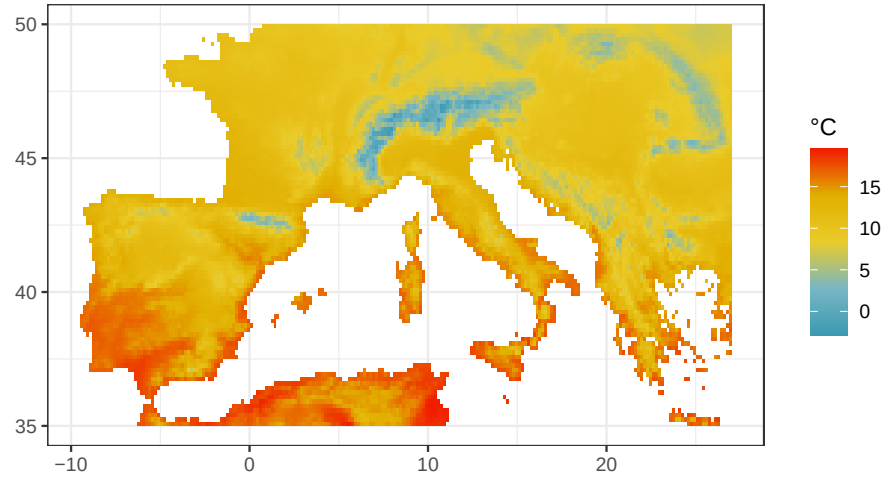
2.2.3.3 Predictor Variables

Monthly variables taken from HadCM3B were temperature (T), precipitation (P), pressure at mean sea-level (SLP), specific humidity (Q) and wind velocity in U and V directions (Wu and Wv). Wu and Wv were not considered as predictor variables in statistical models, but are used to derive two topographic variables described below.

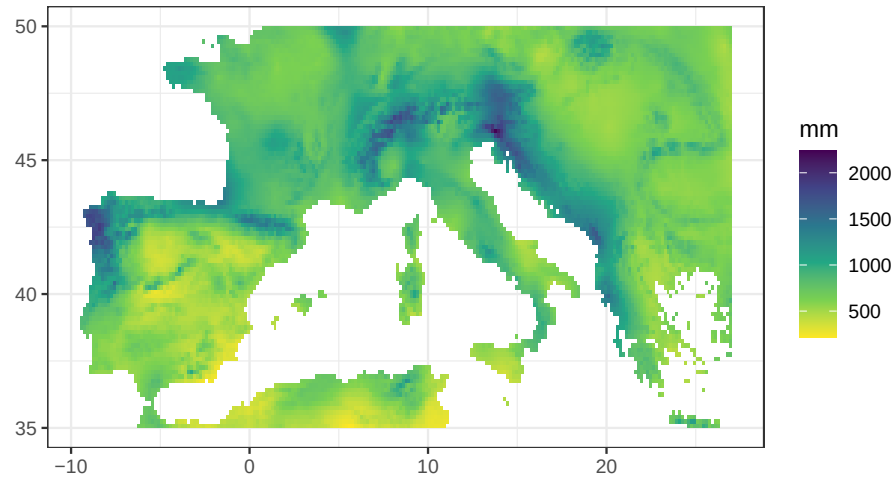
Correlation was initially assessed between all predictor variables, with Q removed given its close correlation with T ($r(134398) = 0.88$, $p < 0.05$). GCM surfaces were interpolated to match the resolution of the response variables using

¹In addition to GAMs, several other model types were trained and tested using the `caret` toolbox in R (Random Forest, Multivariate Adaptive Regression Spline, K-Nearest Neighbours, amongst others). Generally, all models had comparative performance, with only slight differences in RMSE. Given the precedence for using GAMs in palaeoenvironmental downscaling, along with their relatively straightforward characteristics and interpretability, GAMs were pursued in model building and tuning.

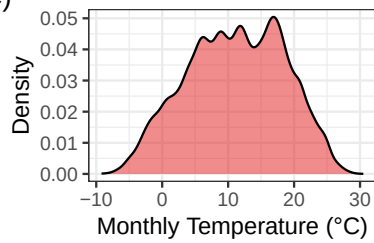
A)



B)



C)



D)

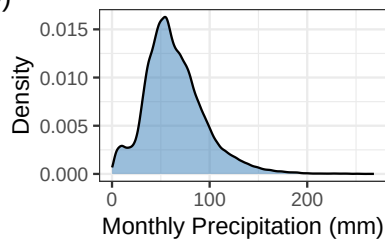


Figure 2.1: Response variables over the study region. Panel A shows mean monthly temperature observations, and Panel B total annual precipitation (New et al. 2002). Panels C and D show probability density functions for mean monthly temperature and precipitation variables, respectively, over all months and gridcells. Temperature shows an approximately Gaussian distribution, whilst precipitation shows a Gamma distribution.

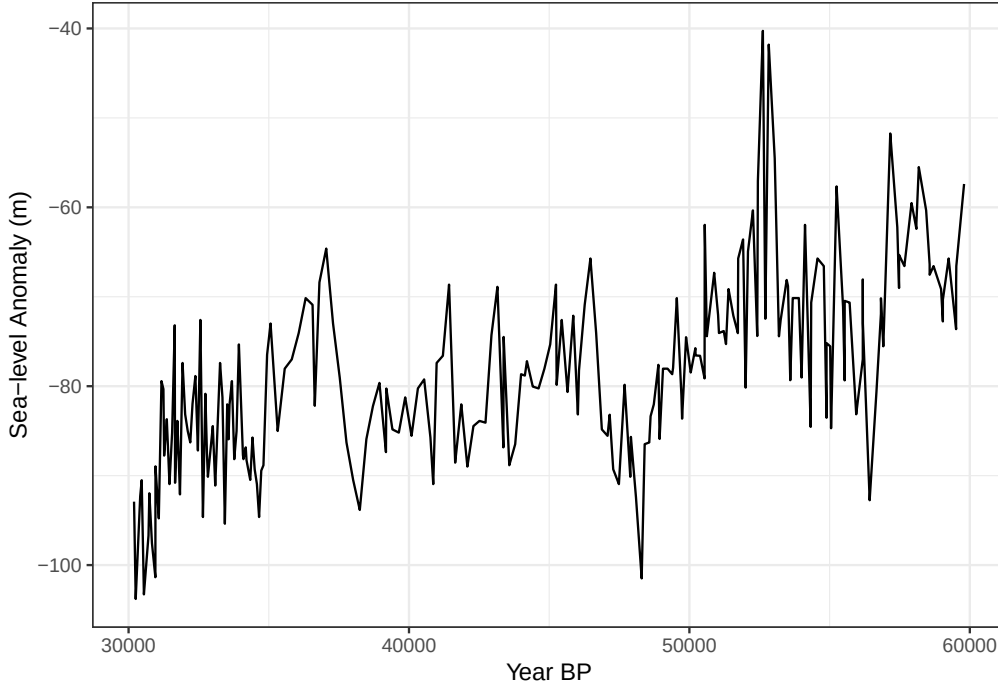


Figure 2.2: Sea-level curve used to adjust the SRTM-30 Plus DEM, reconstructed from the Red Sea sediment core LC21 (Grant et al. 2012).

bicubic interpolation. 100-year means for all GCM variables were taken from the interpolated and corrected HadCM3B time steps between 60 and 30.1 kyr BP. The most recent (0 ky BP) GCM time step was considered a present-day simulation from which to draw predictors for model building.

Four topographic variables were considered as potential predictors, all derived from the SRTM-30 Plus dataset (Becker et al., 2009). This digital elevation model (DEM) is a combination of satellite and bathymetric measurements, from which sea-level and coastline changes between 60 and 30 kyr BP were reconstructed. SRTM-30 Plus was aggregated to match the resolution of the CRU dataset (0.16°), and adjusted by the continuous and high-resolution sea level curve shown in Figure 2.2 (Grant et al., 2012). Elevation (Z) was taken directly from the aggregated DEM. Diffusive continentality (Dco) measures the distance from the centre of each grid cell to the nearest coastline (Vrac et al., 2007).

Following Vrac et al. (2007), advective continentality (Aco) was derived to account for the process by which an air mass becomes increasingly “continental” as it moves over land from ocean (Levavasseur et al., 2011; Vrac et al., 2007;

Wuillez et al., 2014). For each grid cell, Aco is approximated by a sigmoid function applied to a series of radial paths, that originate from the centre of the grid cell and terminate at the paths' intersections with the coast. These paths represent the possible directions from which an air mass might travel to the gridcell's location. Integrating over each path's sigmoid function gives the respective path's measure of continentality, C . The value for Aco is given by the average of C over all radial paths, weighted by the dot product of the mean (prevailing) wind direction over the study region (given by Wu and Wv) and the path directions vector. In contrast to topographic variables Z and Dco , Aco varies by month, as it is dependent on Wu and Wv .

The final topographic variable was derived to simulate rain shadow (Rs) effects and the impact of mountainous regions on local precipitation patterns, building on similar concepts to those of Aco . Using Wu and Wv from the GCM, a path (w) is drawn from the centre of a given grid cell at a trajectory (θ), in opposition to the wind direction ($\theta_w = \arctan(\frac{Wu}{Wv})$). As for Aco , w represents the path of an incoming air parcel to that location. The path is followed (backwards) to an adjacent grid cell, where the trajectory is adjusted to the new grid cell's θ . This process is repeated until the path length reaches 250 km, or the path intersects with the coast. Rs is given by the integral of the elevation profile (Z) along the path:

$$Rs = \int_0^{250} Z dw \quad (2.6)$$

Table 2.1, Figure 2.1, and Figure 2.3 summarise the response and predictor variables used in model building.

2.2.3.4 Spatial Cross-validation and Feature Selection

The `caret` Machine Learning toolbox was used for model building, tuning, and cross-validation, written in R (R Core Team, 2020). Data was split 70:30 for training and testing, respectively, based on random sampling from stratified percentile groupings of each response variable. The test dataset was held out from all model building steps described below and was used for performance evaluation only.

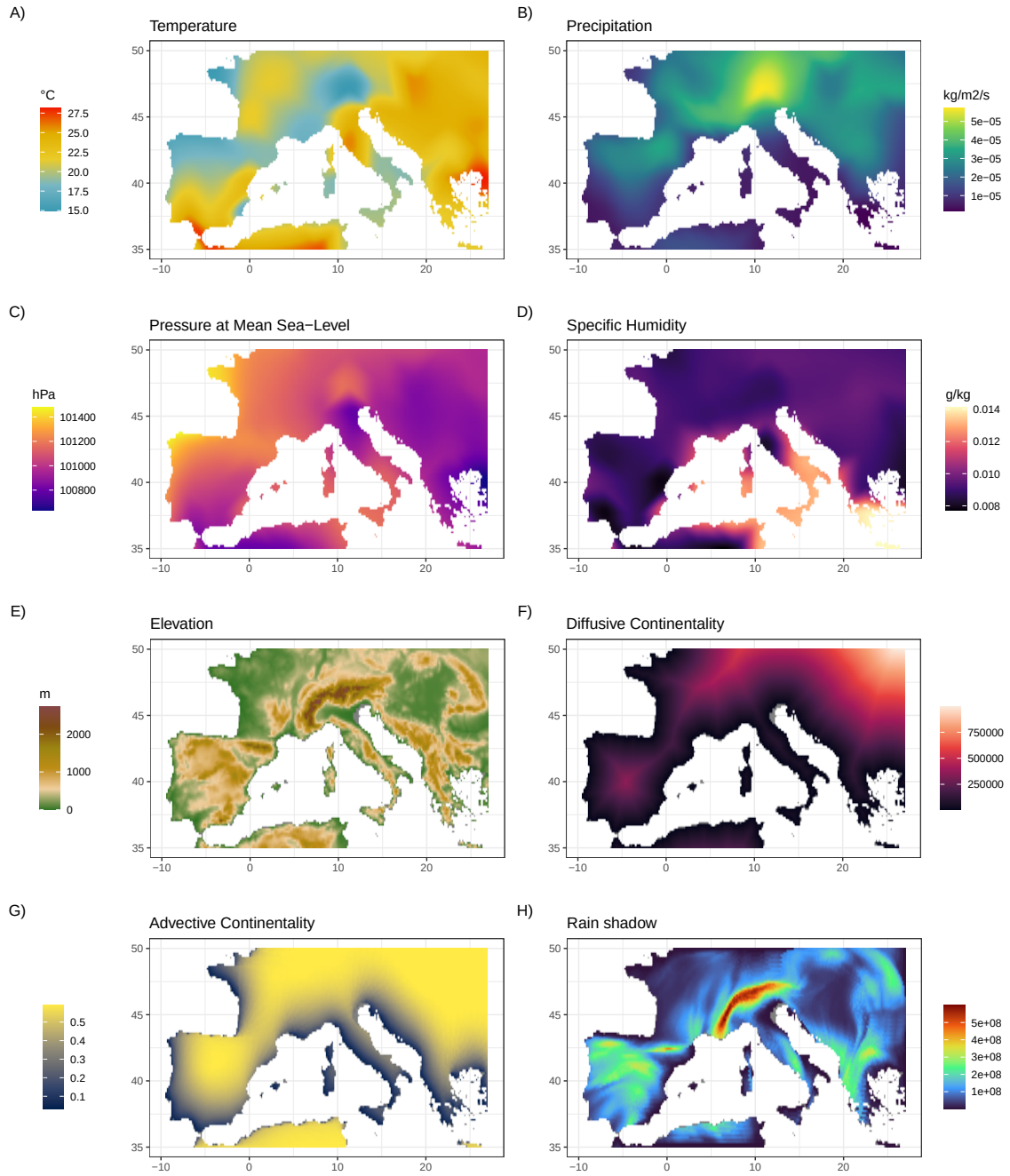


Table 2.1: Predictor and response variables used in statistical downscaling.

Variable	Notation	Type	Source	Resolution	Reference
Temperature	T_{obs}	Response	CRU	0.16°	New et al. (2002)
Precipitation	P_{obs}	Response	CRU	0.16°	New et al. (2002)
GCM Wind in U Direction	W_u		HadCM3	3.75° x 2.5°	Armstrong et al. (2019)
GCM Wind in V Direction	W_v		HadCM3	3.75° x 2.5°	Armstrong et al. (2019)
GCM Temperature	T	Predictor	HadCM3	3.75° x 2.5°	Armstrong et al. (2019)
GCM Precipitation	P	Predictor	HadCM3	3.75° x 2.5°	Armstrong et al. (2019)
GCM Sea-level Pressure	SLP	Predictor	HadCM3	3.75° x 2.5°	Armstrong et al. (2019)
GCM Specific Humidity	Q	Predictor	HadCM3	3.75° x 2.5°	Armstrong et al. (2019)
Elevation	Z	Predictor	SRTM-30 Plus	0.01°	Becker et al. (2009)
Diffusive Continentality	Dco	Predictor	Derived	0.16°	Vrac et al. (2007)
Advective Continentality	Aco	Predictor	Derived	0.16°	Vrac et al. (2007)
Rain Shadow	Rs	Predictor	Derived	0.16°	This study

Before all feature selection and cross-validation, predictor variables were centred and scaled in `caret`, by subtracting the mean of each predictor variable then dividing by its standard deviation. A Box-Cox transformation was also applied to predictor variables (Box & Cox, 1964).

Spatial k -fold cross-validation was used in feature selection methods. This approach has shown to be effective in accounting for spatial autocorrelation in spatial statistical problems (Meyer et al., 2019; Roberts et al., 2017). Here, the “blocking” approach of Roberts et al. (2017) was applied, using a k -means algorithm to split the training data into 20 spatial folds, clustered by their longitude and latitude values. When using randomised cross-validation with spatial data, it is very possible (and probable) that data points selected for training are very similar to those that are tested on, resulting in an artificially lower root-mean square error (RMSE) or higher R^2 value (Meyer et al., 2019). Spatial cross-validation helps to alleviate this problem and assess model bias more readily.

For feature selection, GAMs were trained for all possible combinations of features. After Q was removed due to its correlation with T , seven candidate features were available for feature selection, giving a total of 120 possible combinations of

features, with a minimum of two features for each model. For each combination, cross-validation was implemented by iteratively removing spatial folds and training on the remaining data points, using the removed fold as test dataset. This allows for comparison of models by the mean and standard deviation of their errors across folds, giving some indication of the distribution of model performance over the study region, as well as any spatial biases in the models.

The final model and selected combination of features for each response variable was chosen using the so-called “one standard error rule” (OSE) (Breiman et al., 2017; M. Kuhn, 2016). Here, the simplest model (in this case, the model with the least number of features) that has an RMSE within one standard error of that of the best overall scoring model is selected for final model training. Standard errors were calculated across spatial folds as described above. In the case of a tie, the model with the lowest RMSE error was chosen. Bayesian Information Criterion (BIC) was also calculated for comparison for each feature combination for comparison, a feature selection method frequently used in statistical downscaling (Latombe et al., 2018; Vrac et al., 2007).

The final selected models for each variable were tested on the “hold-out” test data, separated before feature selection and cross-validation. As for the training data, the test data was split into spatial folds using a k-means algorithm in order to infer the spatial distribution of errors. All model performance metrics reported in Section 2.3 are for the test data only, unless stated.

2.2.4 Köppen-Geiger Scheme

Downscaled precipitation and temperature predictions were synthesised in a Köppen-Geiger climate classification scheme to condense results to a cohesive and interpretable output (Beck et al., 2018; Peel et al., 2007). The Köppen-Geiger classification is widely applied in climatology and ecology, centred on a 3-level categorisation scheme which takes monthly temperature and precipitation values for a given location as input. The updated scheme as described by Willmes et al. (2017) was applied, as detailed in Table 2.2.

Table 2.2: Köppen-Geiger climate classification scheme.

First	Second	Third	Description	Criteria
A			Tropical	$T_{min} \geq 18$
	f		Rainforest	$P_{min} \geq 60$
	m		Monsoon	Not (Af) and $P_{min} \geq 100 - \frac{\bar{P}}{25}$
	w		Savannah	Not (Af) and $P_{min} < 100 - \frac{\bar{P}}{25}$
B			Arid	$\bar{P} < 10P_{thr}$
	w		Desert	$\bar{P} < 5P_{thr}$
	s		Steppe	$\bar{P} \geq 5P_{thr}$
		h	Hot	$\bar{T} \geq 18$
		k	Cold	$\bar{T} < 18$
C			Temperate	$T_{max} > 10$ and $0 < T_{min} < 18$
	s		Dry Summer	$P_{minS} < 40$ and $P_{minS} < \frac{P_{maxW}}{3}$
	w		Dry Winter	$P_{minW} < \frac{P_{maxS}}{10}$
	f		No Dry Season	Not (Cs) or (Cw)
		a	Hot Summer	$T_{max} \geq 22$
		b	Warm Summer	Not (a) and $T_{m10} \geq 4$
		c	Cold Summer	Not (a) or (b) and $1 \leq T_{m10} < 4$
D			Cold	$T_{max} > 10$ and $T_{min} \leq 0$
	s		Dry Summer	$P_{minS} < 40$ and $P_{minS} < \frac{P_{maxW}}{3}$
	w		Dry Winter	$P_{minW} < \frac{P_{maxS}}{10}$
	f		No Dry Season	Not (Ds) or (Dw)
		a	Hot Summer	$T_{max} \geq 22$
		b	Warm Summer	Not (a) and $T_{m10} \geq 4$
		c	Cold Summer	Not (a), (b) or (d)
		d	Very Cold Winter	Not (a) or (b) and $T_{min} < -38$
E			Polar	$T_{max} < 10$
	t		Tundra	$T_{max} > 0$
	f		Frost	$T_{max} \leq 0$

$\bar{T}$ = mean annual temperature

T_{min} = mean temperature of the coldest month

T_{max} = mean temperature of the warmest month

T_{m10} = number of months where temperature is above 10 °C

$\bar{P}$ = mean annual precipitation

P_{min} = precipitation of the driest month

P_{minS} = precipitation of the driest summer month

P_{minW} = precipitation of the driest winter month

P_{maxW} = precipitaiton of the wettest winter month

P_{thr} varies, see Peel et al. (2007).

Table 2.3: Palaeoenvironmental records used for validating downscaling predictions over MIS 3. Column "Resolution" refers to the approximate chronological resolution per sample in years, as estimated by Fletcher et al. (2010).

Site	Country	Measurement	Source	Latitude	Longitude	Resolution	Reference
Ejulve	Spain	$\delta^{18}\text{O}$	Speleothem	40.76	-0.59	50	Perez-Mejias et al. (2019)
Hölloch	Germany	$\delta^{18}\text{O}$	Speleothem	47.38	10.15	10	Moseley et al. (2014)
Kleegruben	Austria	$\delta^{18}\text{O}$	Speleothem	47.09	11.67	10	Spötl et al. (2002)
Pozzo Cucù	Italy	$\delta^{18}\text{O}$	Speleothem	40.90	17.16	100	Columbu et al. (2020)
Villars Cave	France	$\delta^{18}\text{O}$	Speleothem	45.30	0.50	100	Genty et al. (2010)
Abric Romani	Spain	Pollen	Sediment	41.53	1.68	200	Burjachs et al. (2012)
Ioannina	Greece	Pollen	Lake	39.75	20.85	350	Tzedakis et al. (2002)
Lagaccione	Italy	Pollen	Lake	42.57	11.80	420	Sanchez Goñi et al. (2017)
Megali Limni	Greece	Pollen	Lake	39.10	26.32	150	Sanchez Goñi et al. (2017)
Monticchio	Italy	Pollen	Lake	40.93	15.62	210	Allen et al. (1999)
Tenaghi Philippon	Greece	Pollen	Lake	40.97	24.22		Wulf et al. (2019)
Valle di Castiglione	Italy	Pollen	Lake	41.88	12.77	440	Sanchez Goñi et al. (2017)

2.2.5 MIS 3 Downscaling and Validation

Variables were derived as described above for MIS 3, using [Armstrong et al. \(2019\)](#)'s dataset and SRTM-30 Plus surfaces adjusted by [Grant et al. \(2012\)](#)'s sea level curve. Using the models selected by the cross-validation and feature selection strategy described in Section 2.2.3.4, high-resolution surfaces were produced for temperature and precipitation variables at 100-year time steps between 60 and 30 kyr BP.

Model output was compared to several published speleothem and pollen records, summarised in Table 2.3 and Figure 2.5. Locations of these records are shown in Figure 2.4. Although many other European records exist for MIS 3 (e.g., [Fletcher et al., 2010](#)), these were selected on the basis of their relatively high-resolution (< 500 years per sample) and their coverage of MIS 3.

All records were condensed to single signals: for speleothem records, $\delta^{18}\text{O}$ measurements were used ([Columbu et al., 2020](#); [Genty et al., 2010](#); [Moseley et al., 2014](#); [Pérez-Mejías et al., 2019](#); [Spötl & Mangini, 2002](#)); and for pollen records, the percentage of pollen species representing the temperate forest biome was used ([Allen et al., 1999](#); [Burjachs et al., 2012](#); [Sánchez Goñi et al., 2017](#); [Tzedakis et al., 2002](#)), other than at Tenaghi Philippon, where the percentage arboreal pollen was used ([Staff et al., 2019](#); [Wulf et al., 2018](#)). Modelled and observed data were

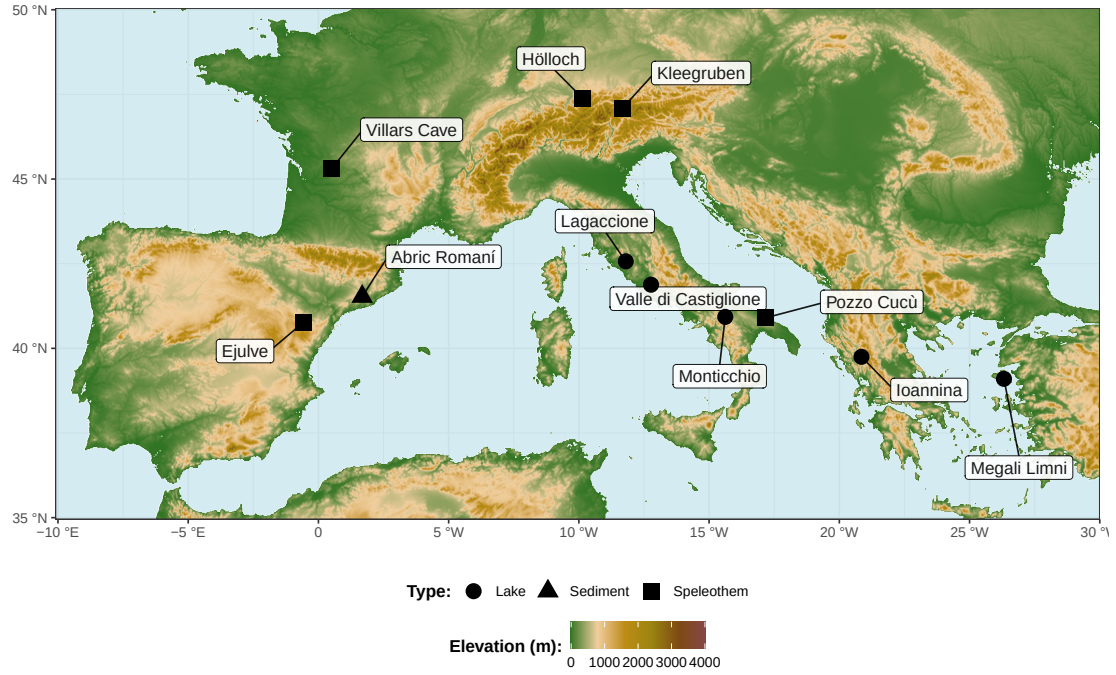


Figure 2.4: Locations of palaeoenvironmental records used to validate GAM predictions.

de-trended by subtracting a least-squares linear function from their time series, before calculating the Pearson correlation coefficients between them. To account for varying resolutions and sampling rates in observations, modelled predictions were approximated at time steps of the validation data by linear interpolation.

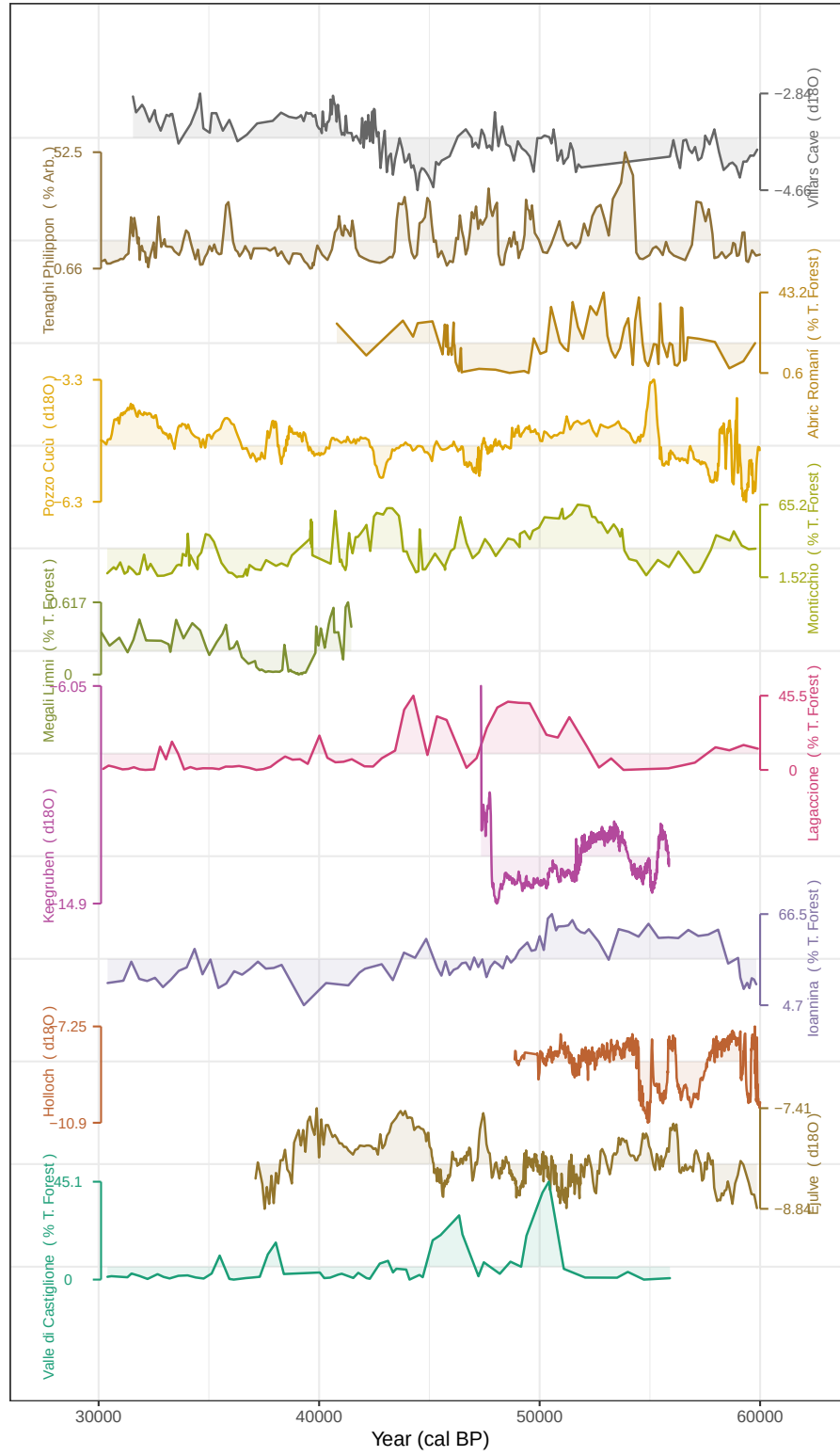


Figure 2.5: Records used to validate downscaled predictions for MIS 3. References for validation records are provided in Table 2.3.

Table 2.4: Feature selection results for Temperature (T_{obs}) and Precipitation (P_{obs}) response variables. Column "N" gives the number of features for each model. The first row shows results for models built using all features, as selected by BIC. For rows below, the three models with the lowest RMSE for each number of features are shown. BIC and RMSE for each feature combination are shown, with σ_{RMSE} and SE calculated across spatial folds. RMSE for GAMs with features selected by OSE are shown in bold.

N	Temperature						Precipitation					
	Features	BIC	RMSE	σ_{RMSE}	SE		Features	BIC	RMSE	σ_{RMSE}	SE	
7	<i>T, P, SLP, Z, Dco, Aco, Rs</i>	392982	1.78	0.49	0.109		<i>T, P, SLP, Z, Dco, Aco, Rs</i>	865439	19.43	5.23	1.170	
6	<i>T, P, SLP, Z, Dco, Rs</i>	395413	1.8	0.49	0.110		<i>T, P, Z, Dco, Aco, Rs</i>	867460	19.49	5.40	1.207	
	<i>T, P, SLP, Z, Aco, Rs</i>	397940	1.82	0.53	0.118		<i>T, P, SLP, Z, Dco, Aco</i>	867896	19.51	5.13	1.147	
	<i>T, P, SLP, Z, Dco, Aco</i>	398484	1.83	0.48	0.108		<i>T, P, SLP, Z, Dco, Rs</i>	867914	19.68	5.25	1.174	
	<i>T, P, SLP, Z, Dco</i>	400309	1.85	0.49	0.109		<i>T, P, Z, Dco, Aco</i>	869795	19.59	5.30	1.186	
5	<i>T, P, SLP, Z, Rs</i>	402082	1.86	0.54	0.121		<i>T, P, Z, Aco, Rs</i>	870313	19.72	5.42	1.212	
	<i>T, P, Z, Dco, Rs</i>	403214	1.88	0.47	0.105		<i>T, P, Z, Dco, Rs</i>	869696	19.76	5.44	1.215	
	<i>T, P, Z, Dco</i>	407381	1.92	0.46	0.104		<i>T, P, Z, Dco</i>	871928	19.86	5.33	1.192	
	<i>T, P, SLP, Z</i>	408802	1.93	0.54	0.120		<i>T, P, Z, Aco</i>	872939	19.96	5.22	1.168	
4	<i>T, P, Z, Rs</i>	411237	1.95	0.53	0.117		<i>T, P, Z, Rs</i>	876045	20.35	5.61	1.254	
	<i>T, P, Z</i>	416953	2.01	0.51	0.114		<i>T, P, Z</i>	879210	20.72	5.46	1.221	
	<i>T, Z, Dco</i>	422472	2.08	0.49	0.110		<i>T, P, Dco</i>	877697	20.87	5.24	1.171	
	<i>T, SLP, Z</i>	427509	2.12	0.56	0.126		<i>T, P, Aco</i>	879331	20.94	5.24	1.172	
2	<i>T, Z</i>	433487	2.2	0.52	0.117		<i>T, P</i>	884187	21.61	5.29	1.184	
	<i>T, P</i>	450969	2.34	0.60	0.133		<i>P, Aco</i>	889384	21.86	5.44	1.216	
	<i>T, SLP</i>	459260	2.45	0.65	0.146		<i>P, Dco</i>	891947	22.09	5.55	1.240	

2.3 Results

2.3.1 Feature Selection

Feature selection by spatial cross-validation using BIC and OSE is summarised in Table 2.4. Feature selection by BIC resulted in all possible features being selected for both precipitation and temperature models. The improvement in model performance when features were selected by BIC over models with features selected by OSE was marginal. For temperature, selected features by OSE were T , P , SLP , Z and Dco , all with p values < 0.001 . For precipitation, selected features were T , P , Z and Dco , all with p values < 0.001 .

Compared quantile-quantile (QQ) plots for BIC and OSE feature selection methods and their modelled residuals on unseen test data are shown in Figure 2.6. For temperature, residuals are approximately normally distributed for both models with some indication of long tails, particularly for extreme negative values.

Scatter plots of residuals against fitted values by BIC and OSE models, also in Figure 2.6, show a random distribution of errors for temperature.

For precipitation, residuals for unseen data are approximately normally distributed for BIC but with very heavy positive and negative tails. Scatter plots of residuals against fitted values for precipitation show clear heteroskedasticity for both BIC and OSE models, with feature selection by OSE resulting in slightly less heteroskedasticity than by BIC. QQ plots for precipitation shows very heavy-tailed distributions of errors (Figure 2.6C), with feature selection by OSE a positive bias, perhaps indicating systematic over-prediction for extreme values.

Figure 2.7 shows RMSE and σ_{RMSE} calculated across spatial folds for each month, generally showing a marginal improvement in predictive performance by BIC over OSE, but greater standard deviations of errors for some months (e.g., November precipitation). Models were most accurate for April and May, with overall poorer performance and larger standard deviations across folds for winter months.

All further reporting of results below relates to GAMs with features selected by OSE, as described in Table 2.4.

2.3.2 GAM Performance

The spatial distributions of residuals for both models are shown in Figure 2.8. The GAM for temperature predicts with relatively low error over the study region (RMSE = 1.85°C, $R^2 = 0.94$). This is broadly consistent with findings in previous downscaling studies over Europe (Latombe et al., 2018; Vrac et al., 2007). The distribution of residuals over the study region reveal some spatial biases in models. Consistently negative residuals are shown southern areas of present-day Italy, Corsica, Sardinia, southern Greece, and some specific coastal areas of Iberia and north-eastern France (Figure 2.8).

A near-linear relationship is identified between CRU temperatures and T except for the highest values of T , with an inverse linear relationship for Z (Figure 2.9). Features SLP , P and Rs show non-linear relationships with CRU temperatures

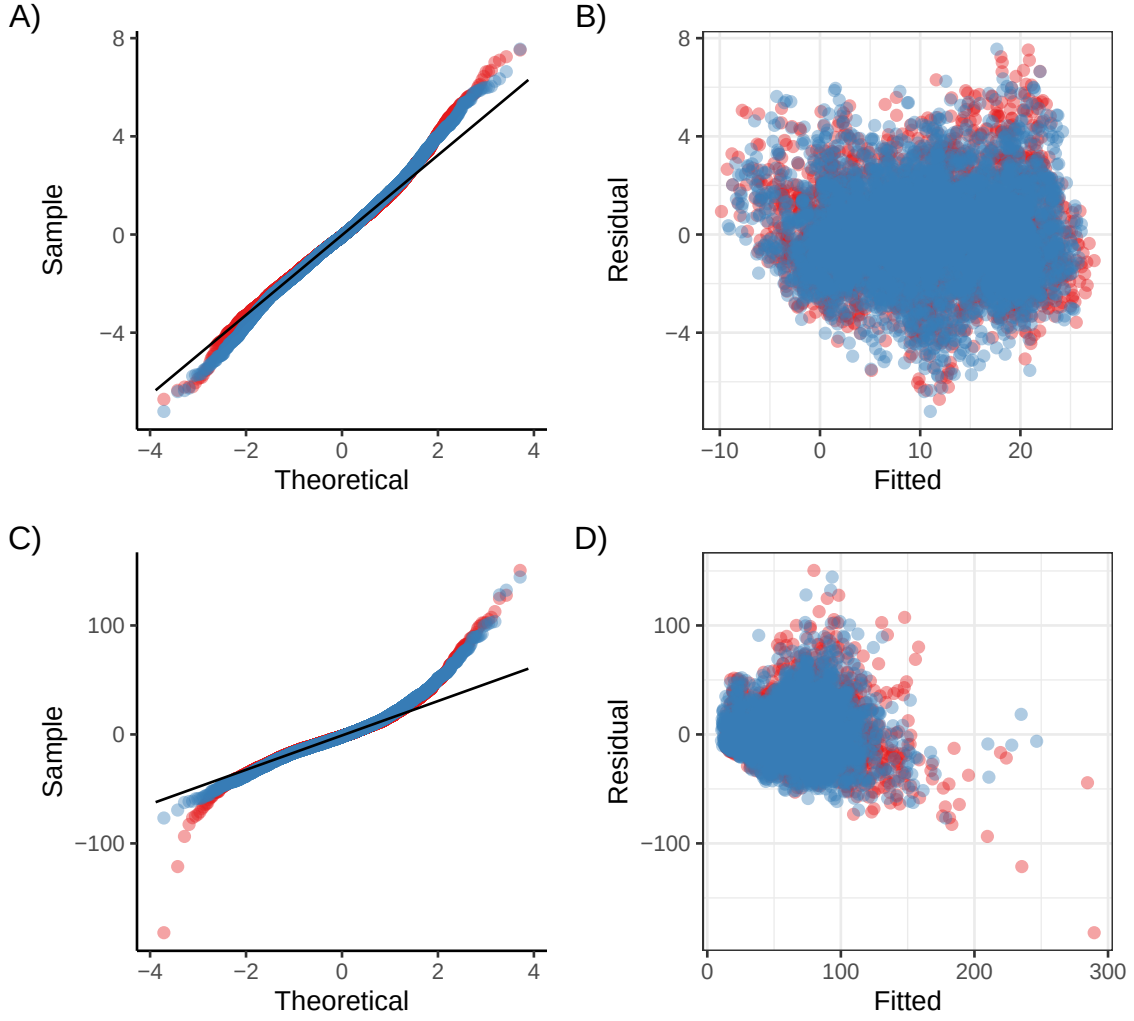


Figure 2.6: Quantile-quantile plots and scatter plots of residuals against fitted values for Temperature (Panels A and B, respectively) and Precipitation (Panels C and D, respectively), for OSE (blue dots) and BIC (red dots) feature selection methods.

but have much lower significance in the GAM than T or Z ; GAM link functions (i.e., $f(X_i)$ in Equation (2.5)) are close to 0 for most values of SLP , for example. Splines for T suggest the GCM simulates monthly temperature variability relatively well at continental scales but over-predicts at higher temperature observations (Figure 2.9, Figure 2.7).

The model for precipitation, by comparison, has a considerably lower proportion of variance explained and relatively high RMSE over the study region (RMSE = 19.86 mm, $R^2 = 0.47$, Figure 2.8). Splines show non-linear relationships between CRU precipitation and T and P , a non-linear relationship close to 0 for larger values of Dco . Z shows a close-to-linear relationship with precipitation (Figure

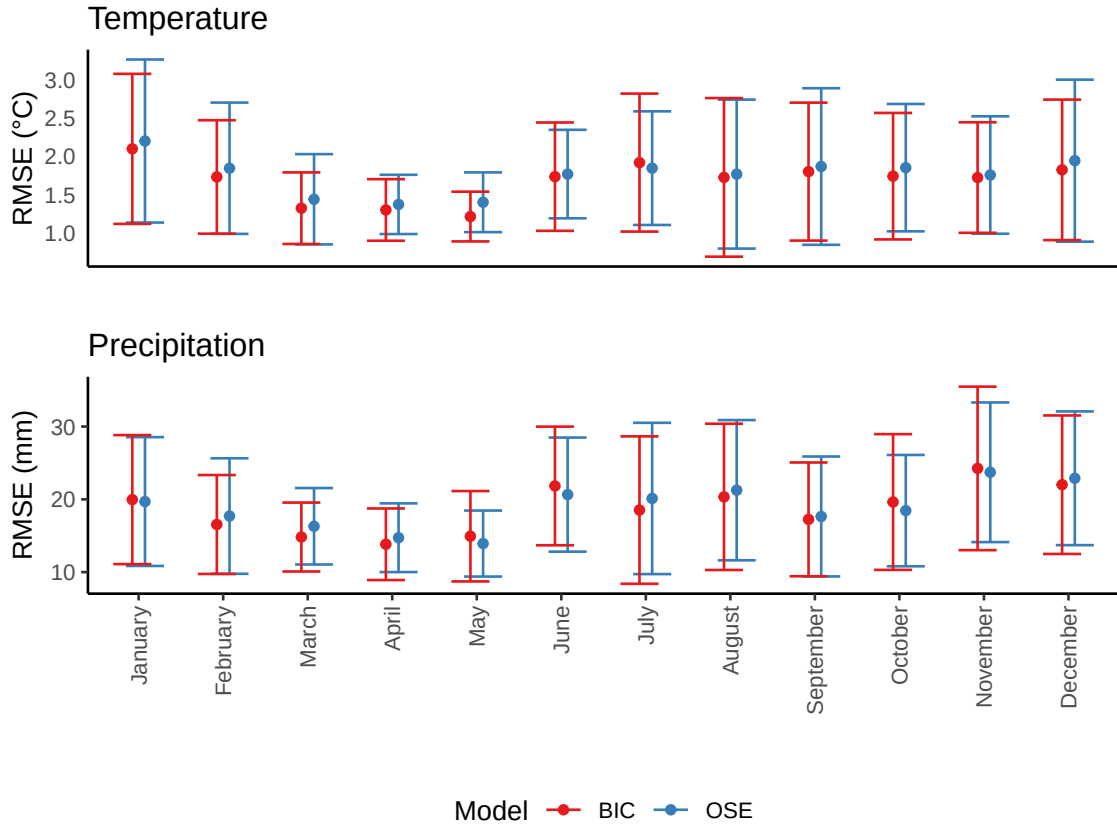


Figure 2.7: Monthly distribution of root mean squared error (RMSE) for Bayesian-Information-Criteria (BIC, red) and one standard error rule (OSE, blue) feature selection methods. Vertical error bars indicate one standard deviation above and below the mean, calculated across spatial folds.

2.10). Spatially, the highest residuals are concentrated over eastern coastal areas of the Adriatic Sea in present-day Greece, Albania, Montenegro, Croatia and Slovenia, as well as over areas of the Pyrenees. As for temperature, the highest RMSE by fold is over southern Italy (Figure 2.8).

2.3.3 MIS 3 Downscaling

Mean, minimum and maximum temperatures and mean total annual precipitation trends over the study region closely resemble those of the GCM (Armstrong et al., 2019), showing that GAMs are strongly dependent on GCM features (Figure 2.12). In turn, GAM predictions also reflect the NGRIP record, given that NGRIP was used to tune the HadCM3B GCM time series (Figure 2.12E). Clear patterns of DO cycles are visible for averaged values over the whole study region. The standard

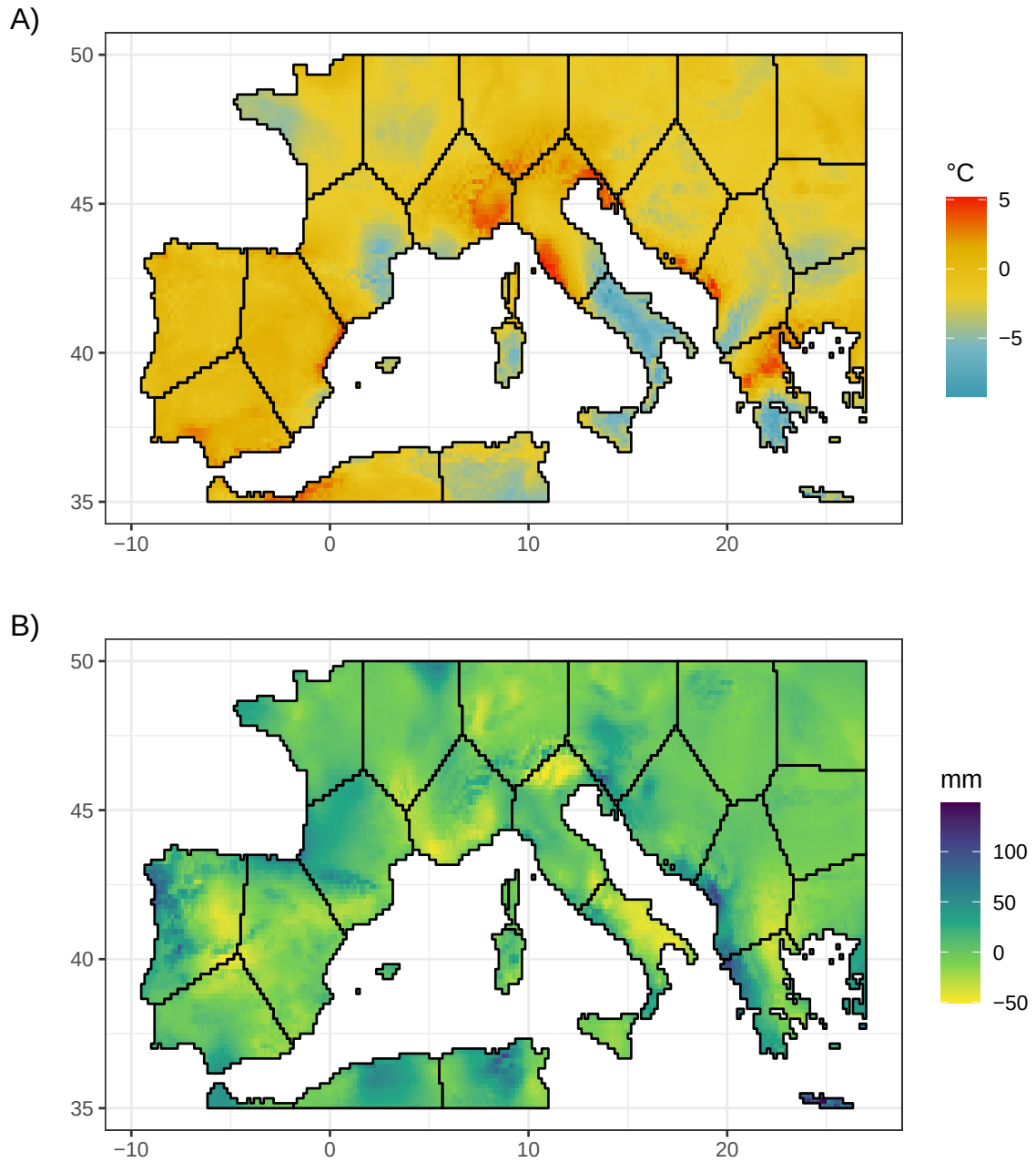


Figure 2.8: Spatial distribution of residuals (observations minus fitted values) over study region for January. Panel A shows temperature residuals, panel B precipitation. Polygons show spatial folds used in cross-validation.

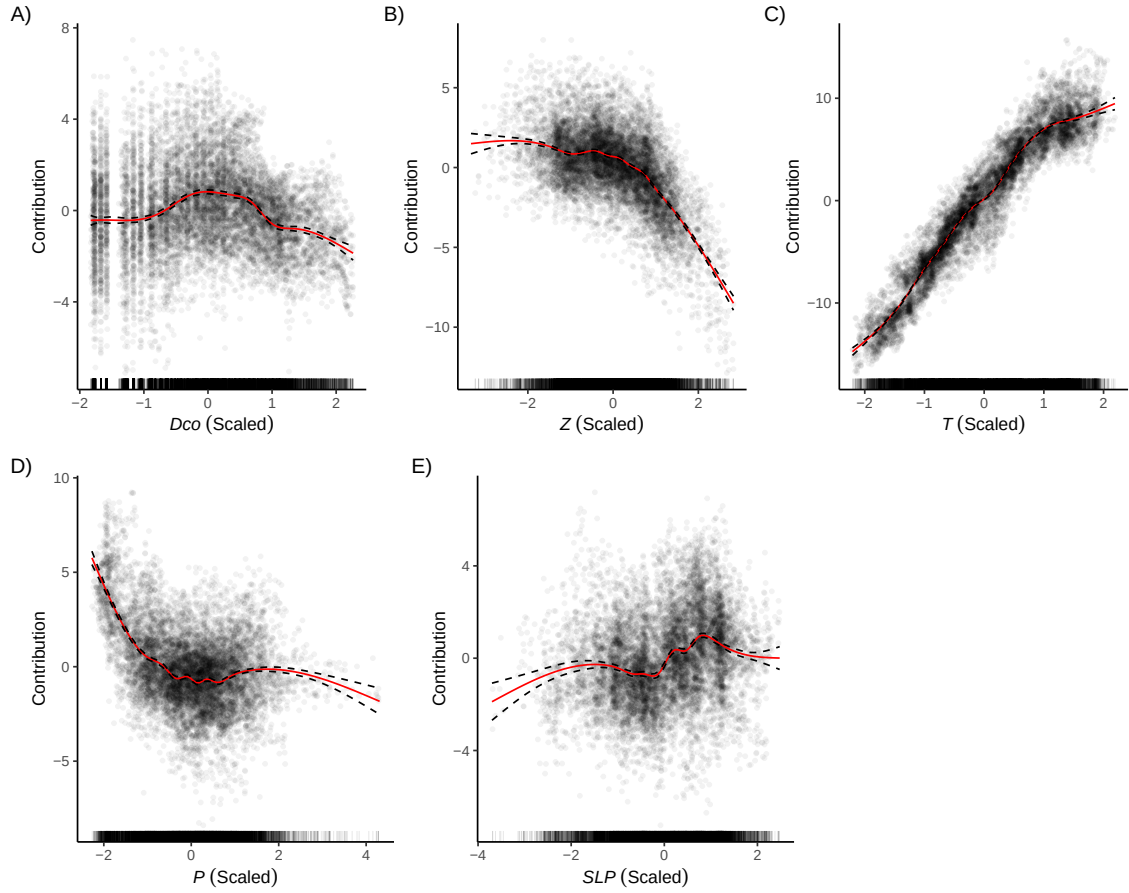


Figure 2.9: GAM link functions fitted to predictors for temperature.

deviation and coefficient of variation for temperature and precipitation, respectively and across months, show an inverse relationship with temperature, indicating an increase in seasonality during colder periods. Greater variability over MIS 3 is clear for winter temperatures and summer precipitation than for their seasonal opposites.

A summary time series for the Köppen-Geiger classification scheme shows inversely related expansions and contractions of temperate (Group C) and continental (Group D) climates (Figure 2.14), reflecting the temperature and precipitation oscillations over DO cycles.

Although DO trends are identifiable across the whole study region, slightly greater variability and amplitude of DO climate fluctuations are seen in western areas. Again, this is driven by the GCM time series (more specifically, the correction surface applied in [Armstrong et al. \(2019\)](#)), rather than being a feature of the GAMs. This pattern is true for both temperature and precipitation predictions, and

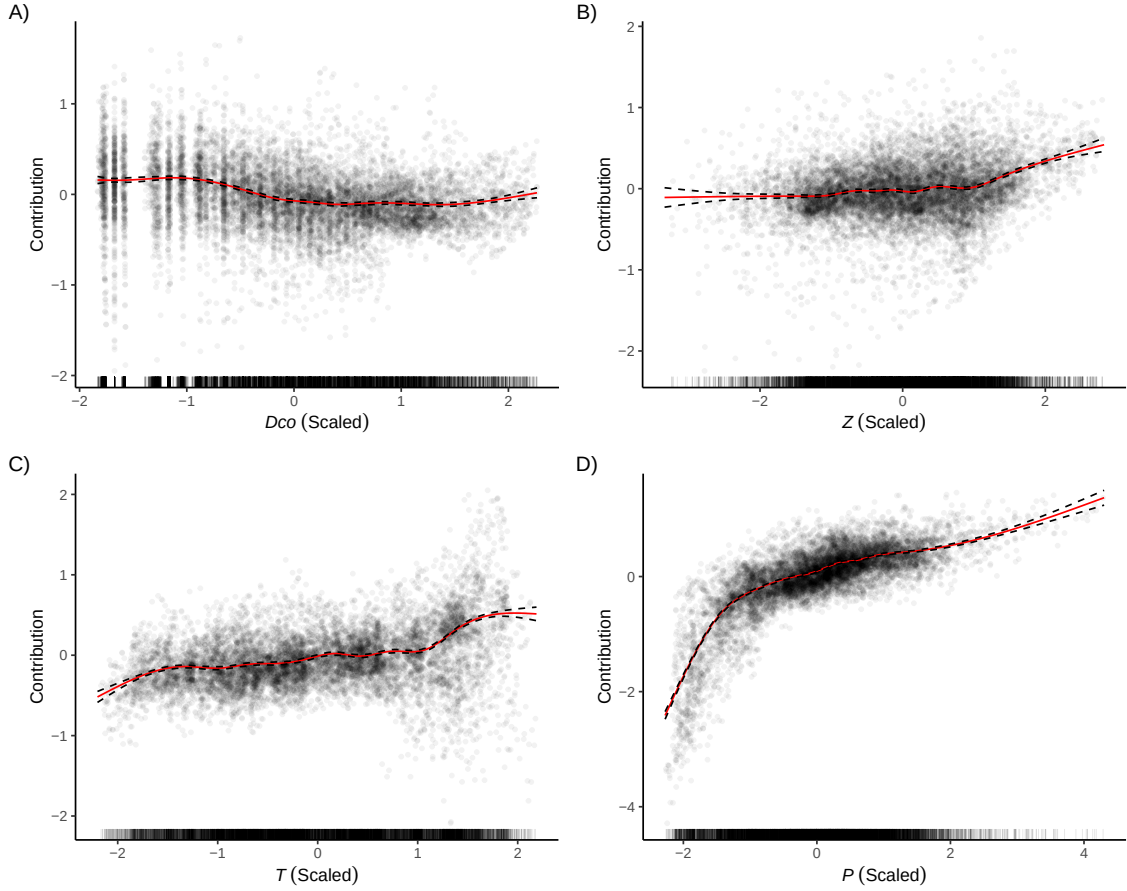


Figure 2.10: GAM link functions fitted to predictors for precipitation.

is particularly evident by inspection of the spatial distribution of the standard deviation of predictions over time steps (Figure 2.13A and 2.13B). Similarity between temperature and precipitation variability is qualified by the spatial distribution of correlation coefficients of mean annual temperatures and total precipitation, which very closely relate for large areas of the study region ($r > 0.95$, Figure 2.13.c).

Temperature and precipitation anomalies, along with Köppen-Geiger classifications for time steps corresponding to GI 12 and GS 12 are shown in Figure 2.15. These time steps are extreme maxima and minima of temperatures when averaged over the study region (46.7 and 43.7 kyr BP, respectively). Climate conditions for GI 12 are predicted as being relatively close to current conditions in Europe, with temperatures on average 1.3 °C lower than CRU temperatures. Much heavier rainfall in western Iberia and warmer temperatures for some regions relative to the present day are predicted for GI 12, with Köppen-Geiger classifications very

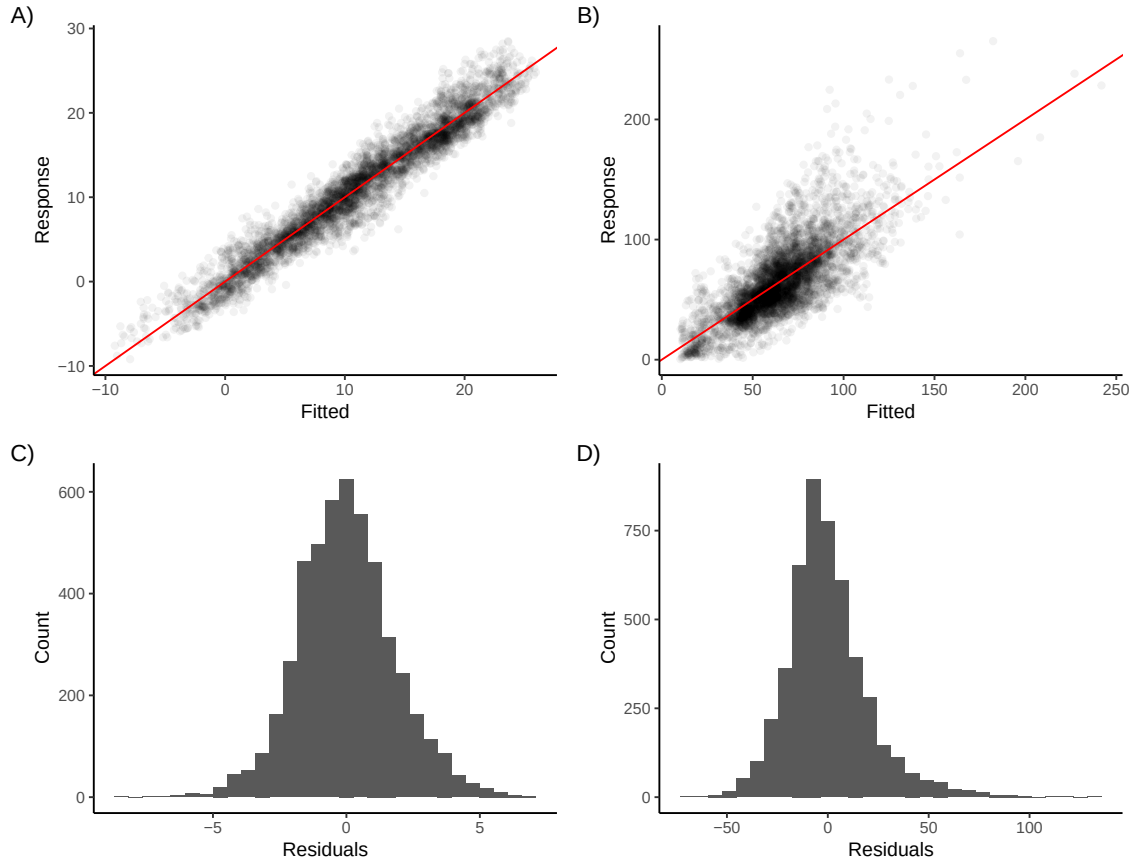


Figure 2.11: Response vs. fitted values and histograms of residuals for temperature (Panels A and C, respectively) and precipitation (Panels B and D).

similar to their modern distribution. This contrasts severely with GS 12, 3 kyr later, with temperatures on average 7.2 °C colder than current conditions and extreme aridity, with expansions of cold semi-arid (BSH) and cold-desert (BWk) regions in southern Iberia and areas of eastern Europe (Figure 2.14).

2.3.4 MIS 3 Validation

Comparison of results to available MIS 3 palaeoenvironmental proxies shows very poor agreement between predictions and records (Figure 2.16). Given that predictions for all locations closely follow the NGRIP record, the records with the closest similarities to GAM predictions (highest correlation coefficients) are those that have previously been highlighted as reflecting the NGRIP record, such as Hölloch and Klee gruben speleothems. At Pozzo Cucù, by contrast, where speleothem records have recently been argued to represent climatic stability through the Middle

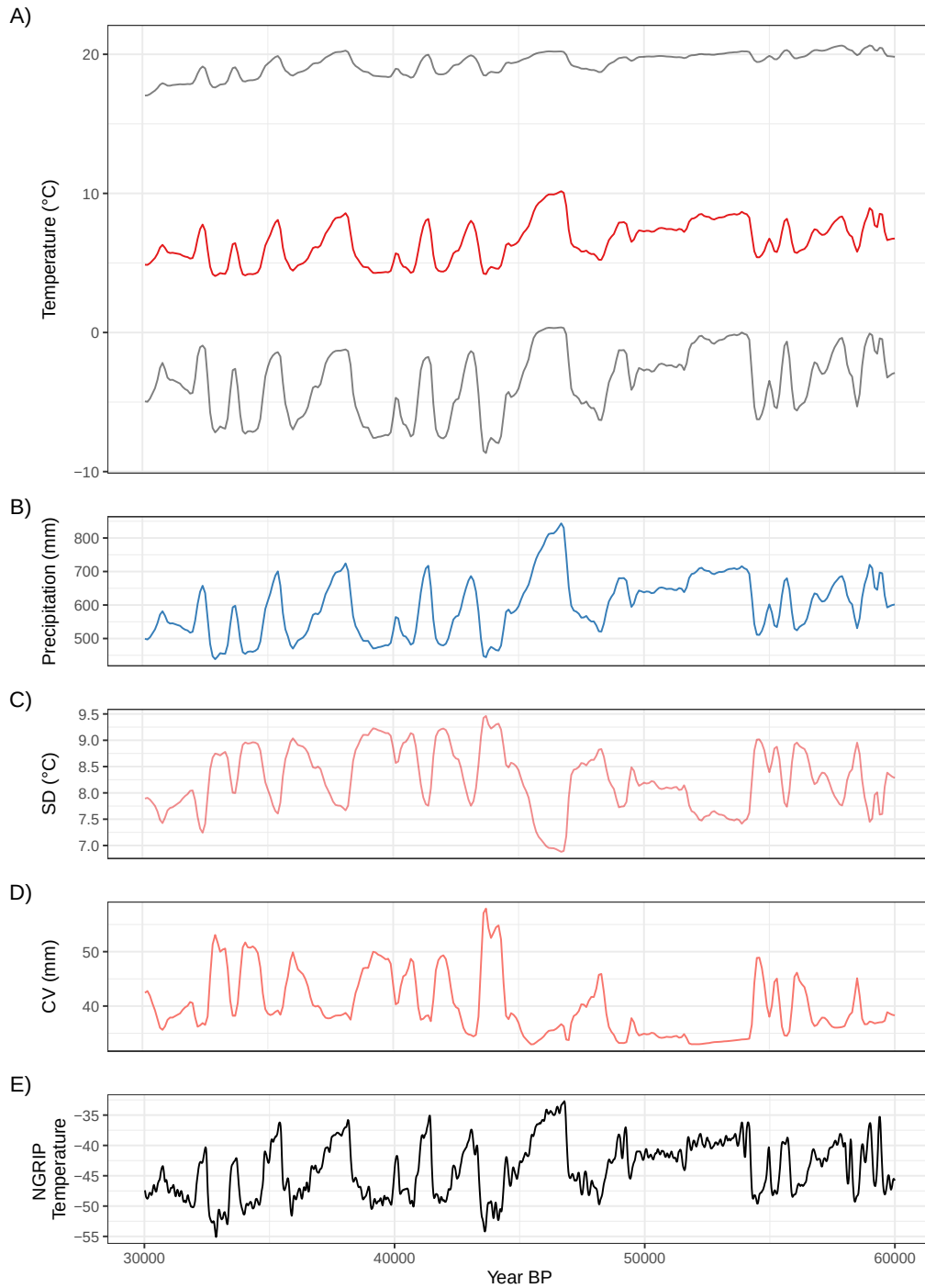


Figure 2.12: Mean annual temperature and annual precipitation predicted by GAMs for the study region over MIS 3. Panel A shows mean annual temperature (red line) with minimum and maximum temperatures above and below, respectively (grey lines). Panel B shows total annual precipitation averaged over the study region. Panel C shows the standard deviation of monthly temperatures, giving some indication of seasonality. Panel D shows the coefficient of variation of monthly total precipitation. Panel E shows NGRIP temperature reconstructions for comparison (Kindler et al. 2014).

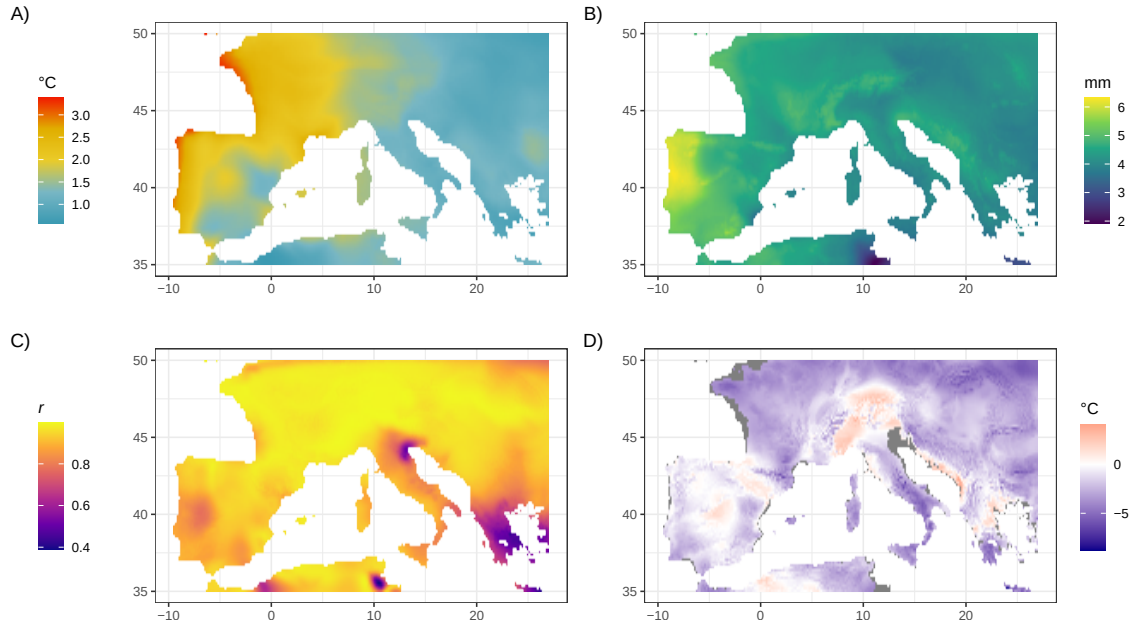


Figure 2.13: Spatial trends over MIS 3. Panels A and B show the standard deviation of mean annual temperatures over MIS 3 and the log of the standard deviation of total annual precipitation over MIS 3, respectively. Greater variability is shown for both temperature and precipitation over western areas of the study region. Panel C shows the correlation coefficient of mean annual temperatures and total annual precipitation for MIS 3. Panel D shows the mean difference in annual temperatures between GAM predictions and the delta method, suggesting overall colder conditions predicted by the delta method.

to Upper Palaeolithic transition, GAM predictions show pronounced variability reflecting DO cycles. Some records, such as Valle di Castiglione and Lagazzione, are clearly too low in resolution for periods of MIS 3 for adequate comparison (Figure 2.5 and Figure 2.4). The closest correlations were between downscaled predictions and the Klee gruben speleothem and Ioannina pollen record.

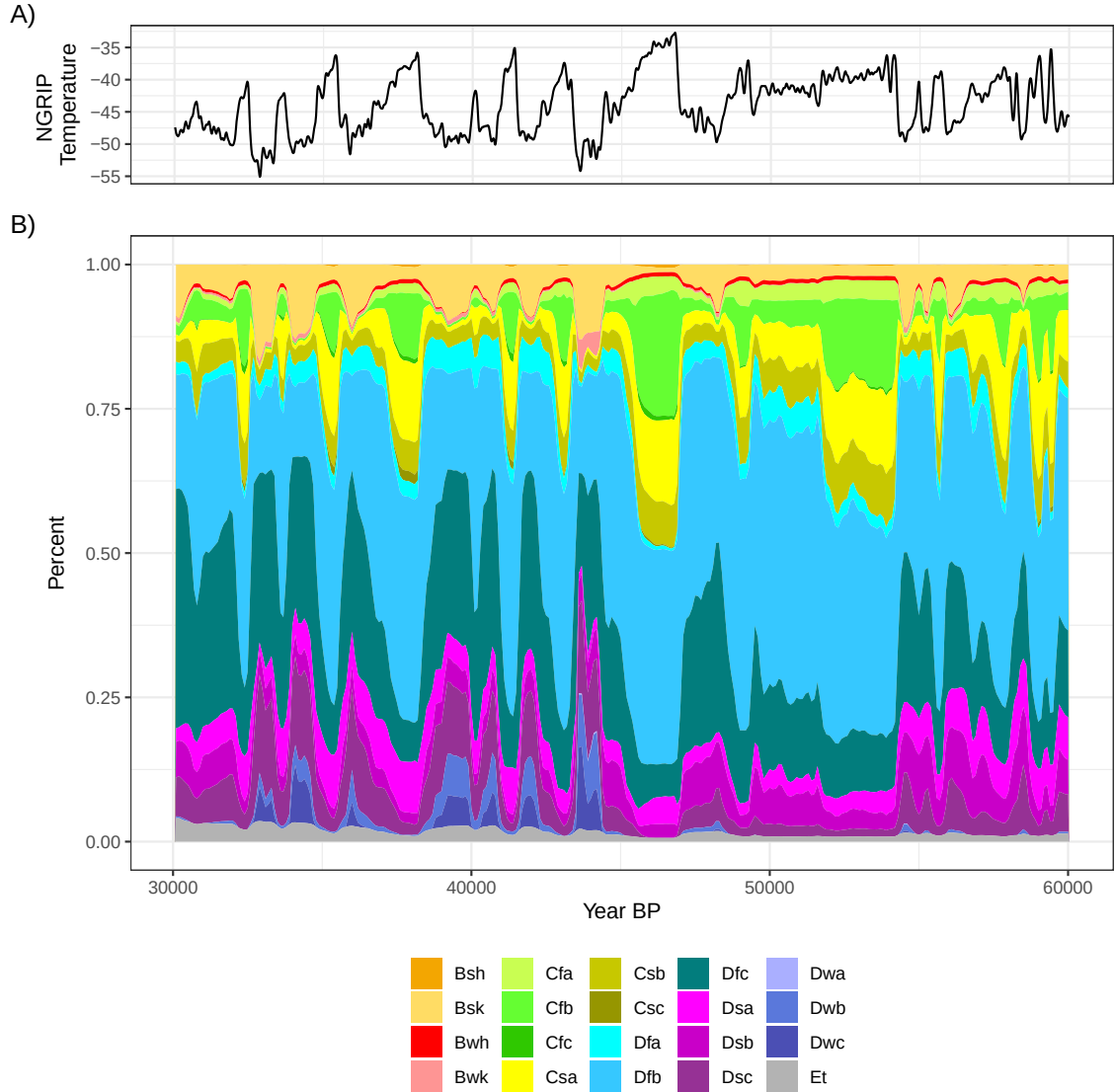


Figure 2.14: Köppen-Geiger classification over MIS 3, based on GAM predictions. Panel A shows NGRIP temperature reconstructions (Kindler et al., 2014). Panel B shows the percentage cover for each climate classification over MIS 3. Expansion and contraction of Arid (B), Temperate (C), and Cold (D) categories are particularly apparent, corresponding to DO cycles as shown in Panel A. Further description of the Köppen-Geiger classification scheme and its categories is given in Table 2.2, following Willmes et al. (2017).

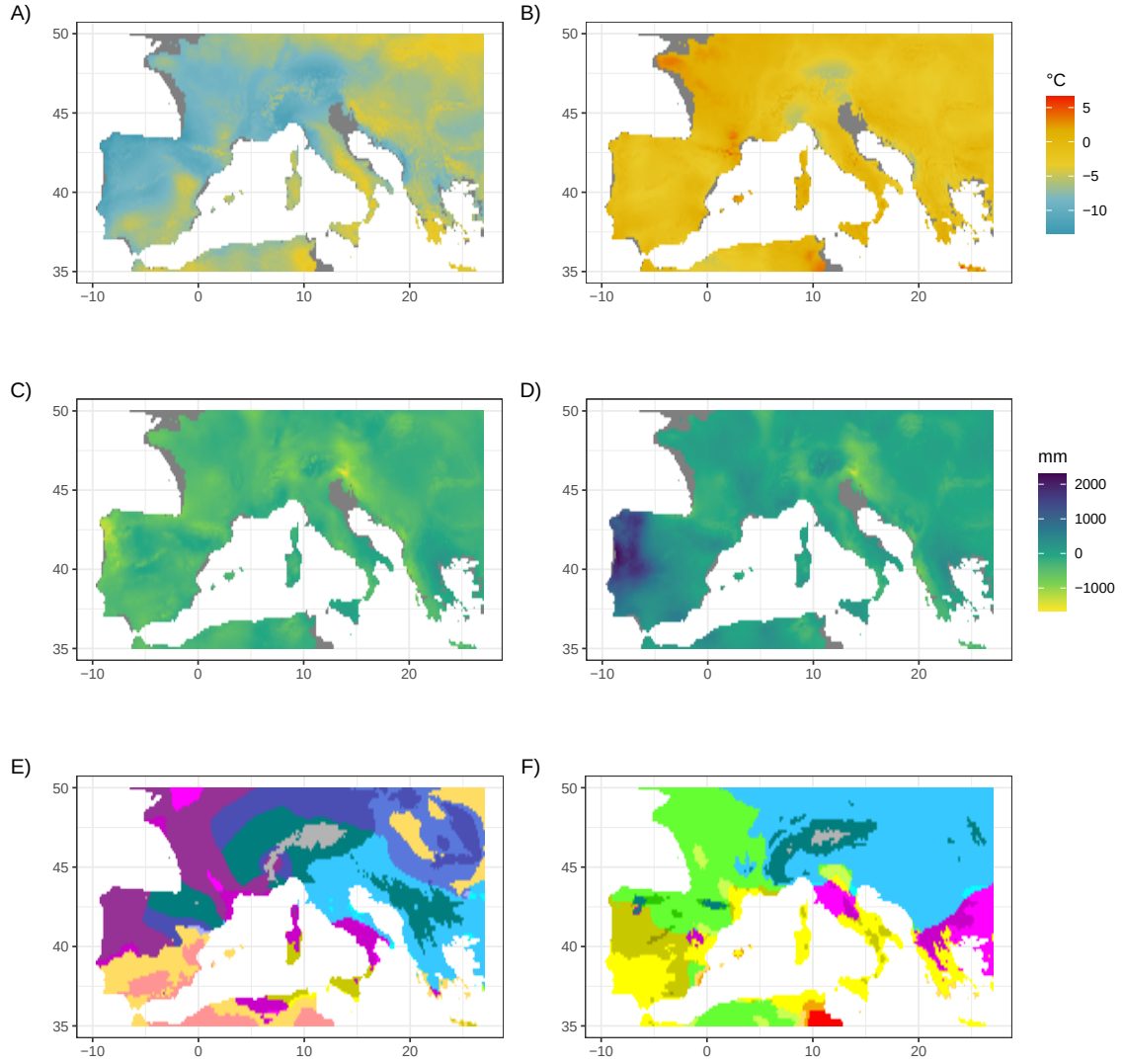


Figure 2.15: Temperature and precipitation anomalies (CRU minus GAM predictions) for 43.7 kyr BP (Panels A and C) and 46.7 kyr BP (Panels B and D). Panels E and F show Köppen-Geiger classifications over the study region for 43.7 and 46.7 kyr BP, respectively.

2.4 Discussion

In this study a continuous GCM time series of climate variability during MIS 3 was downscaled to 0.16° resolution over Europe, using the GAM method for both temperature and precipitation response variables (Vrac et al., 2007). Four GCM variables and four topographic variables were considered as predictor variables. A Machine Learning pipeline was applied to the GAM calibration process, using spatial cross-validation in feature selection, with further validation on “hold-out”

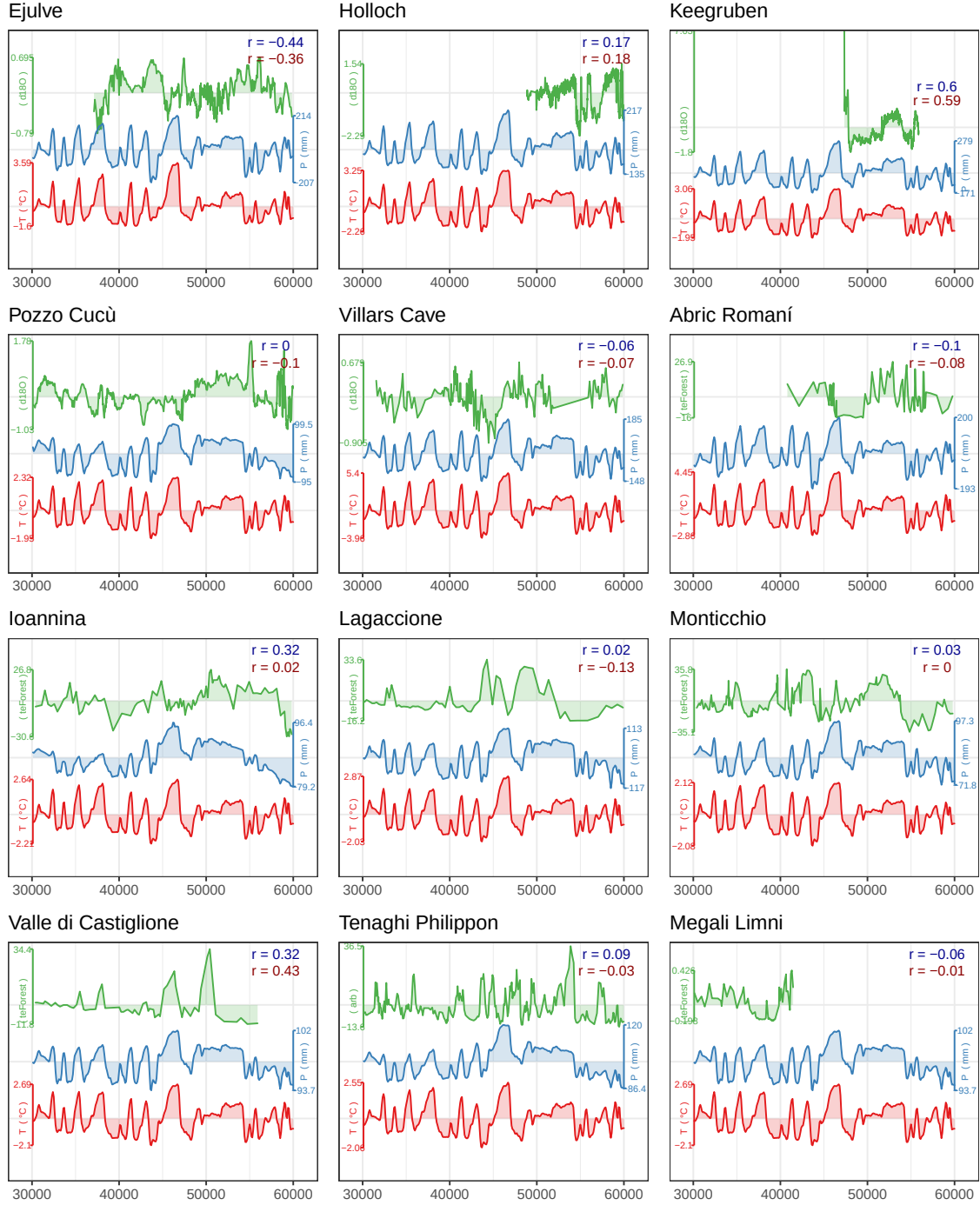


Figure 2.16: Comparisons of palaeoenvironmental proxies (green) and temperature (red) and precipitation (blue) predictions produced by GCM downscaling at their respective locations. Correlation coefficients are shown at the top right corner of each panel, for temperature (red) and precipitation (blue).

data. Results show that temperature is predicted with relatively low error, but precipitation predictions were far less accurate, with particularly heavy-tailed positive residuals and clear heteroskedasticity in their distribution. Residuals for both variables across the study region reveal spatial artefacts and biases in predictions. For precipitation, these results indicate that GCM variables were not sufficient in their spatial resolution for use as predictors, whilst topographic predictors forwarded for feature selection were not sufficient to model climate variability at the resolution of the CRU dataset.

Downscaled predictions for MIS 3 closely follow the $\delta^{18}\text{O}$ pattern of the NGRIP record (Rasmussen et al., 2014), and corresponding temperature reconstructions by Kindler et al. (2014). This is unsurprising, given that the original GCM data was corrected to the NGRIP temperature reconstructions (Armstrong et al., 2019; Kindler et al., 2014), and that the GAMs were heavily dependent on the GCM variables as predictors. Increased seasonality for both variables during stadial periods of DO cycles was inferred. For temperature, this variation corresponds to decreasing winter temperatures during stadial periods, whereas for precipitation increased seasonality corresponded to increased summer aridity. Predictions were compared to available palaeoenvironmental records for validation, indicating poor correlation with most records.

The results of this study highlight fundamental uncertainties in statistical downscaling and validation processes. These must be addressed before confident palaeocological or archaeological conclusions can be made from downscaled predictions, or similarly before results are used as inputs into other predictive models, such as species distribution models or eco-cultural niche models (Banks, 2017). Discrepancies between palaeoenvironmental records and modelled predictions further illuminate challenges for the integration and alignment of proxy data. In the following sections, some of the key uncertainties and areas of attention for future work are briefly discussed.

2.4.1 GAM Performance and Spatial Cross-Validation

Results show that feature selection by BIC selected all possible features, with only a marginal increase in predictive accuracy over OSE (Table 2.4). Previous studies exploring GCM downscaling employing BIC have also resulted in all possible features being selected (Latombe et al., 2018). This is also seen for downscaling experiments using Akaike information criterion (AIC), a similar approach that generally adds *less* cost than BIC to models with higher numbers of features (Vrac & Naveau, 2007). It is very likely that GAMs built by BIC were overfitting “noise” to response variables by their additional features, and that these features were not physically relevant. As such, a more stringent feature selection method than AIC or BIC that penalises more complex models (i.e., larger numbers of features) is appropriate for GCM downscaling. The OSE method offers some alleviation to this problem.

Nonetheless, despite resulting in less features than BIC, it is likely that some features selected by OSE were similarly physically uninformative. Many of the the GAM link functions shown in Figure 2.10 and Figure 2.9 are either excessively complex or have little variance in their contribution, while many of the variables themselves do not appear to add significant value to the models and their predictions. For example, predictor SLP in Figure 2.9 shows a GAM link function to T_{obs} near zero for most values, with a complexity that contradicts physical phenomena such as the ideal gas law. Furthermore, a model using just T and Z as predictors for temperature achieves a relatively high R^2 (0.93; RMSE = 2.2 °C; Table 2.4), suggesting that a very high proportion of variance in T_{obs} can be explained by these predictors alone. The temperature lapse rate of the troposphere means that, broadly, air temperature has a close-to-linear relationship with elevation at local scales, while large-scale spatial and monthly patterns in temperature are typically well-simulated by GCMs. These factors go some way to explain why these two features alone can predict temperature with comparable accuracy to all other models. Given that these variables are also supported by physical logic, there might be some argument for *qualitative* feature selection in

GCM downscaling. A more stringent feature selection method, such adjusting the OSE rule to two standard errors, might also be effective.

Many previous downscaling experiments have shown considerably higher error and lower accuracy for precipitation than for temperature, as found in this study. Precipitation is known to be dependent on a number of complex, non-linear, and interacting factors that are difficult to model universally for a given study area, either numerically or statistically (Vrac et al., 2007; Vrac & Naveau, 2007). Poor predictive performance for P_{obs} suggests that none of the features forwarded for feature selection were adequate for relating GCM variables to precipitation observations. Relationships between Dco or Z and P_{obs} in Figure 2.10 show high dispersion at lower values for each variable (x-axes), perhaps explaining the heteroskedasticity evident in Figure 2.6. GAM link functions for Dco , Z , and T (Figure 2.10) are close to zero (i.e., flat) for most of their values; only P shows significant contribution. As such, it may be necessary to derive more complex topographic variables to improve downscaling predictions for precipitation. Unfortunately, Rs failed to represent rain shadow effects sufficiently to be selected by OSE for the precipitation model, but experimenting with path lengths above or below 250 km, or capturing upward gradients in rain shadow paths (w in Equation (2.6)) might improve the predictive capacity of this variable.

k -fold spatial cross-validation helps to avoid autocorrelation in geostatistical modelling (Meyer et al., 2019). For downscaling, spatial cross-validation by folds was found to be a useful method by which to understand spatial biases in models. Errors and their standard deviation across folds illuminated model biases, while allowing for a more robust feature selection method than has previously been applied to statistical downscaling using GAMs (Burke et al., 2017; Latombe et al., 2018; Vrac et al., 2007). For future downscaling studies, the spatial cross-validation scheme as implemented in this study could also be employed to understand and derive more complex topographic variables, or parameter tune variables such as path length in Rs . Increasing the number of folds might also highlight spatial biases in the models at higher spatial resolution.

The distribution of residuals over the study region reveals clear spatial structure to model errors (Figure 2.8). RMSE for folds over southern Italy was systematically higher for all months for both temperature and precipitation. This is because the resolution of HadCM3B is too low to represent these areas of the Italian peninsula, or, in other words, Southern Italy “disappears” within the coarse spatial resolution of HadCM3B. The effect of this for model building is wider dispersion for predictor variables, and poorer model performance (Figure 2.10). The delta method, by contrast, does not suffer from this drawback. A similar effect is identifiable for some coastal areas, where interpolation of the grid cells of HadCM3B extends its oceanic output over terrestrial areas, resulting in higher errors in south-east Iberia, coasts of the Adriatic and for north-west France (Figure 2.8). Results for these areas should be interpreted with caution; understanding the spatial limitations of the GCM and selecting a study area accordingly is an essential step in designing an appropriate statistical downscaling scheme.

Finally, it could also be worthwhile to explore different climate observation datasets as response variables for statistical downscaling. The CRU observations by New et al. (2002) were produced by interpolating irregularly spaced observations as a function of latitude, longitude, and elevation. As such, elevation (Z) is not a wholly independent variable in the GAMs, as it was used - albeit using a separate dataset (Becker et al., 2009) - to build the response variables. Despite this, the New et al. (2002) data is frequently used for response variables in statistical downscaling models that also include elevation and elevation-derived variables as predictors (Latombe et al., 2018; Levavasseur et al., 2011; Vrac et al., 2007; Woillez et al., 2014). This is a difficult challenge to overcome; many of the alternative climate observation datasets use similar strategies in their construction (e.g., Fick & Hijmans, 2017). Reanalysis datasets such as ERA5 (Hersbach et al., 2020), although being lower in spatial resolution (0.25°), offer the advantage of being wholly independent of predictor variables. A comparison of statistical downscaling methods using different observation datasets might illuminate this bias further.

2.4.2 MIS 3 Downscaling and Validation

Significant climatic variability was predicted over Europe for MIS 3 by the downscaling models. Modelled temperature and precipitation predictions averaged over the study region closely followed the NGRIP record (Figure 2.12), with the extremes of this variability shown by the shift from warm and wet conditions at 46.3 kyr BP to cold and arid conditions at 43.7 kyr BP in Figure 2.15. GAMs were dependent on T and P from the GCM. Although topographic variables adjusted by sea level reconstructions by Grant et al. (2012) were also selected as predictors (Dco and Z), they appeared to have little effect on temperature and precipitation predictions. DO oscillations imposed on the GCM surfaces by Armstrong et al. (2019) were clearly driving the downscaled variability over MIS 3. This suggests that, as for the delta surfaces in the delta method, the GAMs were effectively “static” across time steps, with almost all variability in predictions coming from the GCM.

As highlighted by Latombe et al. (2018), it is essential that training data for statistical downscaling adequately captures the range of values expected for the target period. In the case of Latombe et al. (2018)’s study, for example, climate during the LGM in Europe was reconstructed, with training data selected to reflect the colder and drier range of values expected for the period. For stadial periods of MIS 3, palaeoenvironmental records also suggest colder and drier conditions than today prevailed during stadial periods, but much greater variability is apparent across millennia (Fletcher et al., 2010; Kindler et al., 2014).

In this study some GCM predictors demanded extreme extrapolation beyond the training data. GS 12 is the clearest example of this problem (Figure 2.17). Training values for T show much higher values than for the GS 12 time step, while for P this appears less pronounced. Although GAM link functions suggest extrapolation would be linear (Figure 2.9), understanding this effect in more detail is important for future downscaling studies. Experiments could include an expansion of the study area to include a wider range of training values to include areas of Scandinavia, for example, as for Latombe et al. (2018), or analyses of

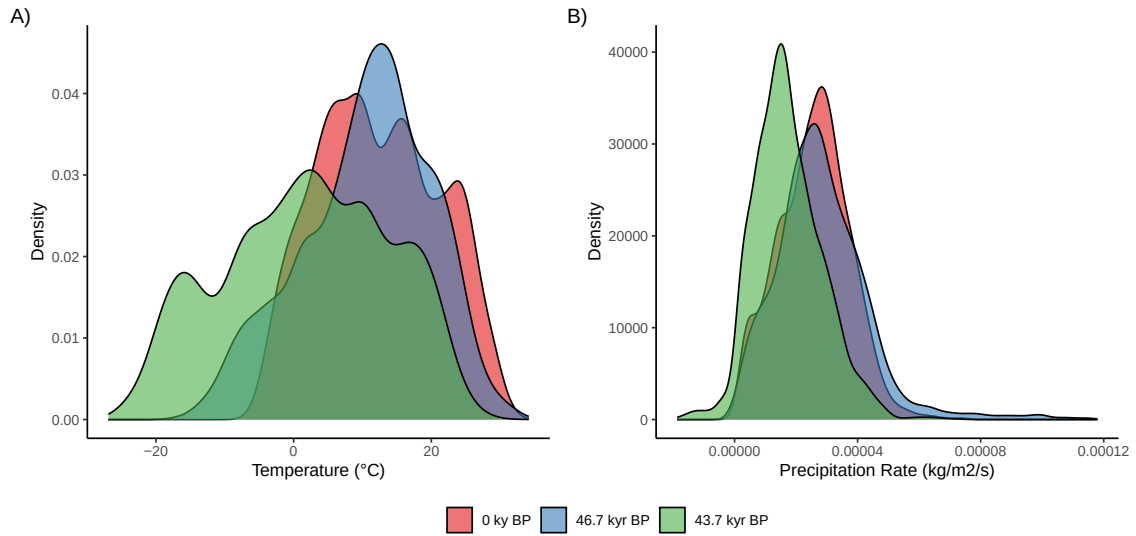


Figure 2.17: Density plots for values of GCM T and P (Panels A and B, respectively), for training data (0 ky BP, red) and two example time steps, GI 12 (46.7 ky BP, blue) and GS 12 (43.7 ky BP, green). T for GS 12 is distributed over much colder temperatures than training data, requiring extrapolation of the GAM function.

the effects of different extrapolation functions in GAMs at extreme periods, such as for Greenland Stadial 12.

The spatial impact of DO events in Europe predicted by downscaling models can be inferred from Figure 2.13A and 2.13B. Greater variability in the west of the study region broadly corresponds with records from the North Atlantic and the confirmed extent of DO events (Blaauw, 2012; Sánchez Goñi et al., 2018). Köppen-Geiger classification variability also agrees with expansion and contraction of woodland environments and xerophytic (semi-desert) species suggested by pollen records at continental scales (Badino et al., 2020). Very high correlation coefficients between temperature and precipitation predictions (> 0.8 , Figure 2.13C) suggest that the two vary in conjunction over MIS 3: colder periods are drier, and warmer periods are wetter.

Comparisons between downscaled temperature and precipitation predictions and key records of palaeoenvironmental change show very poor correlation in all cases except at Klee gruben (Figure 2.16). Many fluctuations in the validation data appear out of phase with modelled predictions, or DO events are missing from validation signals altogether. Speleothem records (Ejulse, Hölloch, Klee gruben,

Pozzo Cucù, and Villars Cave), show high-resolution variability not captured in the 100-year time steps of the modelled predictions, while several pollen records (Lagaccione, Valle di Castiglione) clearly have a resolution too low to allow adequate comparison.

It is beyond the scope of this study to consider why each individual record does or does not show certain DO events, or why records might show oscillations that are out of phase with downscaled predictions (and, by extension, the Greenland record). These questions have been explored elsewhere, either in the original publications of each record ([Allen et al., 1999](#); [Columbu et al., 2020](#); [Tzedakis et al., 2002](#)), or in various reviews of palaeoenvironmental records for MIS 3 ([Badino et al., 2020](#); [Fletcher et al., 2010](#)). Taken collectively, the results of the validation procedure applied in this study highlight some of the more general challenges in model-data comparisons in palaeoenvironmental science, and proxy comparison more broadly. If, as found in this study, model-data comparisons do not agree, it is not necessarily that the model is wrong; there are a collection of possible reasons for discrepancies between modelled climate variables and palaeoenvironmental proxies.

- **Age models**

The key assumption made in this study is that the age models of the validation models represent the true age of the signal, with their respective chronological uncertainties ignored. In practice, this is of course problematic. Obtaining accurate and precise chronological estimates in palaeoenvironmental science can be extremely challenging. For MIS 3, fluctuations in the NGRIP record show variability on centennial scales ([Rasmussen et al., 2014](#)), lower than the uncertainties of most of the chronologies of records used in this study.

Speleothem chronologies might be considered to be relatively well constrained due to their uranium-thorium dates ([Adolphi et al., 2018](#); [Corrick et al., 2020](#)), but periods of hiatus in their growth can mean that key portions of MIS 3 are missing from their signals. Multiple pollen records considered here,

meanwhile, have significant uncertainty attached to their chronologies. Some of these chronologies were produced wholly by manual tuning to other records and linear interpolation, such as at Ioannina, discussed further in Chapter 3. Updated radiocarbon determinations for Tenaghi Philippon, meanwhile, have shown that previous measurements underestimated the age of sections of the core considerably (Müller et al., 2011; Staff et al., 2019). It is perhaps for this reason that Armstrong et al. (2019) predominantly used ice core records for their validation, given that ice core chronologies are generally far better constrained than many of the records considered here.

The approach taken here does not account for any uncertainty in existing age models, or the methods used to produce them, and this is a significant limitation of the validation strategy in this study.

- **“Correlation” as a measure of similarity**

Pearson correlation was used to measure similarity between the modelled predictions and observations. Given the uncertainties associated with the validation chronologies described above, Pearson correlation is not an effective measure of similarity for these signals. Although records were detrended before calculating their correlation, the complex nature of the records and the construction of their chronologies is not accounted for by Pearson correlation.

Alternative similarity metrics that account for variability in a signal’s magnitude and resolution, such as Dynamic Time Warping (see Chapter 3 and Chapter 4) may be more appropriate for palaeoenvironmental model-data comparisons. Or, a qualitative assessment of the relationship between records can be made, as typically given alongside publications of proxy records.

- **Sampling resolution**

As already noted, the validation records show considerable variability in their sampling resolution. Pollen records at Ioannina, Lagaccione, and Valle di Castiglione have relatively low resolution (i.e., > 300 years per sample), while

speleothem records are typically measured at higher resolution than the 100-year time steps of this study. Reconciling the signals of records with varying resolution is a key challenge in palaeoenvironmental science, particularly for MIS 3. High-resolution speleothem records might be filtered to remove local variability, but low-resolution records present a much more difficult problem. It is very difficult to ascertain whether DO-like signals can be inferred at Lagaccione, for example, because the sampling resolution of the record is lower than the duration of some DO cycles.

- **Measures of environmental variability**

Lastly, it should be noted that varying signals taken as proxies in this study ($\delta^{18}\text{O}$, % Temperate Forest pollen, % Arboreal pollen) represent complex and interacting environmental processes. Localised changes in climatic, ecological, or geological conditions can affect their signals, and these nuances are difficult to account for in model-data comparison. Again, if a certain peak, trend, or shape of a modelled signal is not found in observations, one should not necessarily conclude that the model is wrong.

Validation methods clearly require improvement before confident archaeological interpretation can be made from model predictions. Among the possible expansions to validation methods are using directly dated faunal remains at archaeological sites as an environmental proxy and more applying more complex time series comparison algorithms (Burstyn et al., 2021; Lisiecki & Lisiecki, 2002). The latter is discussed in more detail in Chapter 3, where attempts to align the signals of downscaled predictions and three of the validation records considered are made.

2.4.3 Relevance for the Archaeology of the Middle to Upper Palaeolithic Transition

GI 12 and GS 12 are particularly pronounced time steps in the downscaled time series, with an extreme change over a 3 kyr period: from broadly temperate conditions at 46.7 ky BP, to very cold and arid conditions at 43.7 ky BP (Figure 2.15).

At present, the warming pulse at ~ 47 kyr BP coincides with the earliest evidence for *Homo sapiens* in Europe at Bacho Kiro at ~ 46.9 ky calBP (Fewlass et al., 2020). It is also possible that later and repeated oscillations between stadials and interstadials between 44 and 38 kyr BP (GI and GS 12-9) caused repeated hominin population-depopulation cycles (Staubwasser et al., 2018). With reference to the archaeology of the Middle to Upper Palaeolithic transition, Müller et al. (2011) links the preceding deterioration of Heinrich Event 5 at ~ 48 ky calBP at Tenaghi Philippon, Greece, with a “demographic vacuum” in Neanderthal populations, that was subsequently filled by *Homo sapiens* following the rapid expansion of woodland biomes at around 47 ky calBP. Chronologies for the estimates at Tenaghi Philippon have recently be revised (Staff et al., 2019), but the ecological aspects of this hypothesis certainly fit with the modelled results of this study.

However, before drawing conclusions from their predictions, it is essential to acknowledge that comparisons to existing environmental records for MIS 3 showed very poor agreement (Figure 2.16), spatial biases were apparent in both model’s residuals (Figure 2.8), and performance metrics for GAMs for precipitation showed relatively high RMSE and low R^2 values (Table 2.4). Any archaeological inference from the modelled temperature and precipitation estimates should be viewed with extreme caution. Similar uncertainties are often overlooked in applications of downscaled GCM data to archaeological and palaeoecological questions.

A varied range of high-resolution palaeoenvironmental proxies exist for Europe, while GCMs and associated downscaling methods are becoming increasingly sophisticated (Baño-Medina et al., 2020). These broad and complex data sources have excellent potential for exploring the environmental contexts of the Middle to Upper Palaeolithic transition, when used in combination. Although the spatial and chronological resolution of GCM-derived data can be rich, and its potential for subsequent species or eco-cultural niche modelling appealing, there are currently considerable discrepancies between modelled predictions and observations. These discrepancies must be acknowledged and addressed before GCM-derived data is applied to archaeological questions. The challenge of appropriately comparing

palaeoenvironmental simulations to observations to understand the limitations of the former in palaeoenvironmental reconstruction should be prioritised and embraced in future.

2.5 Conclusion

This chapter presented a study in statistical downscaling of a continuous GCM simulation of climate variability during MIS 3 over Europe. Statistical models (GAMs) were built to capture the relationship between climate observations and large-scale atmospheric variables provided by the GCM. Temperature was predicted with high accuracy, but poor predictive performance was found for precipitation. A robust cross-validation and feature selection scheme ensured that model metrics were accurate, revealing spatial biases in the models relating to the coarse resolution of the GCM. Temperature and precipitation were downscaled to high spatial resolution at 100-year time steps between 60 and 30 kyr BP, and compared to European palaeoenvironmental records for validation. The validation process revealed clear discrepancies between modelled and observed data, which could be attributed to poor model performance, and/or a number of reasons relating to the proxy records.

The validation method of this chapter raises a key question in integrating palaeoenvironmental records: if two independently dated proxy records' respective signals do not agree with each other, should we assume that either the proxy signal or underlying age model is incorrect? Or, put another way, if both the sampling resolution and age models of two palaeoenvironmental records from the same region are considered robust and reliable, should we expect to see synchronous patterns in their signals? This problem is explored in more detail in the following chapter, where three high-resolution pollen signals for MIS 3 are aligned to downscaled predictions by numerical methods, and their age models compared to original estimates for the signals' chronology.

3

Synchronised Climate Proxies for Marine Isotope Stage 3 by Dynamic Time Warping

Abstract

Chapter 2 explored the use of downscaled climate model output to reconstruct the environmental conditions of the Middle to Upper Palaeolithic transition in Europe. In attempting to validate downscaled climate model predictions against records of past environmental climate variability, the challenge of assimilating various chronologies of palaeoenvironmental change to a common time scale was highlighted. In this chapter, a potential remedy to this problem is explored.

Dynamic Time Warping (DTW) is a signal processing algorithm that compresses or expands the time-ordered indices of one signal to those of another, based on the Euclidean distance between the signals at all possible offsets of their indices. In this study, DTW is applied to three high-resolution pollen records, employed as proxies for palaeoclimatic change over MIS 3 (60 to 30 kyr BP). The three pollen records are aligned by DTW to the downscaled temperature and precipitation predictions produced in Chapter 2. The resulting realigned age models are compared to

the records' existing chronologies, derived from radiocarbon measurements, event layers, or manual tuning strategies.

Alignment to downscaled predictions by DTW produces realistic age models for each pollen record, but these show discrepancies with their existing chronologies. This calls into question either: a) the reliability of existing chronologies; b) the application of quantitative tuning strategies to environmental records; or c) the assumption that large-scale climatic events were synchronous over MIS 3, and are detectable at the resolution of Late Pleistocene pollen records. These points are discussed in more detail before outlining recommendations for future quantitative tuning approaches.

3.1 Introduction

In Chapter 2, a GCM simulation was statistically downscaled using Generalised Additive Models, to reconstruct climate variability over Europe during MIS 3 (Armstrong et al., 2019; Vrac & Naveau, 2007). Because the GCM surfaces driving the downscaling models were corrected using the Greenland record, the resulting time series for the whole study region closely resembled those of $\delta^{18}\text{O}$ records from Greenland ice cores (Armstrong et al., 2019; Kindler et al., 2014; Rasmussen et al., 2014), but comparisons to terrestrial MIS 3 proxies showed poor agreement. As discussed in Chapter 2, there are many possible reasons for these model-data discrepancies. A significant and problematic assumption of the approach taken in Chapter 2 was that the underlying age models of the validation records represent the true age of the records.

For many palaeoenvironmental records for MIS 3, this is rarely likely to be the case, particularly for pollen records (Fletcher et al., 2010). Significant uncertainties in chronology building for palaeoenvironmental proxies can relate to a scarcity of material for dating, ^{14}C calibration challenges and reservoir effects, radiocarbon contamination (see Chapter 5), or discontinuous and unclear deposition processes such as sediment mixing (Fletcher et al., 2010). Combined, these factors mean that the age models attached to many of the palaeoenvironmental records used for validation in Chapter 2 carry considerable uncertainty.

Pollen records provide some of the highest-resolution terrestrial archives of environmental change over MIS 3, and serve as the basis for many reconstructions of climate variability throughout this period (Allen et al., 1999; Fletcher et al., 2010; Sánchez Goñi et al., 2018; Tzedakis et al., 2002). Pollen counts measured at intervals throughout a sediment core or sequence can be used to infer past climate variability, either by their absolute or proportional values, or by translating their signals to a reconstruction of a given climate variable by transfer functions (Chevalier et al., 2021; Kühl et al., 2002; Marsicek et al., 2018; Salonen et al., 2019; Telford & Birks, 2009). The resulting time series are rarely considered in isolation;

the signals from one record are typically compared others' for the same period, sometimes across wide spatial scales (Fletcher et al., 2010). However, in order to reliably draw parallels between records, such comparisons require that each record is supported by a robust and reliable chronology.

Palaeoenvironmental proxies often invite some qualitative flexibility in their interpretation. As Blaauw notes, “it is well known that [palaeoenvironmental] data are uncertain and thus can be ‘moved around’ in order to fit certain hypotheses” (2012, p. 38). Under the assumption that fluctuations in multiple palaeoenvironmental signals represent large-scale changes to earth system circulation patterns, and therefore were synchronous, records with uncertain chronologies are often aligned to other records to produce an updated chronology. This practice is known as “tuning,” and is common where palaeoenvironmental chronologies are sparse, or when dating is not possible at all (Allen et al., 2000; Follieri et al., 1998; Magri, 1999; Margari et al., 2007; Shackleton et al., 2000; Sirocko et al., 2021).

There are many approaches to tuning, but broadly the process involves aligning the peaks or shape of one palaeoenvironmental signal to those of another, either qualitatively (Shackleton et al., 2000) or quantitatively (Lin et al., 2014; Lisiecki & Lisiecki, 2002; Muschitiello & Aquino-lopez, 2023). Qualitative tuning allows room for nuance in chronology building, but its *ad hoc* nature invites the possibility of subjective bias. Quantitative approaches using numerical methods, meanwhile, can offer a more objective solution to the tuning process, where the signals of palaeoenvironmental records are aligned deterministically by optimisation algorithms. This avoids the problem of subjectivity in visual alignment, but can instil false confidence in the resulting age models, while overlooking the local nuances or “leads and lags” across multiple palaeoenvironmental signals.

Many of the relevant pollen records considered for validation in Chapter 2 were produced by qualitative tuning, combining chronological measurements with alignment using tie points. For example, the chronology of the Ioannina I-284 core was produced by manual alignment for MIS 3, between a single radiocarbon date calibrated to around 25 ky calBP at 23 m depth, and a single tie point at 36 m

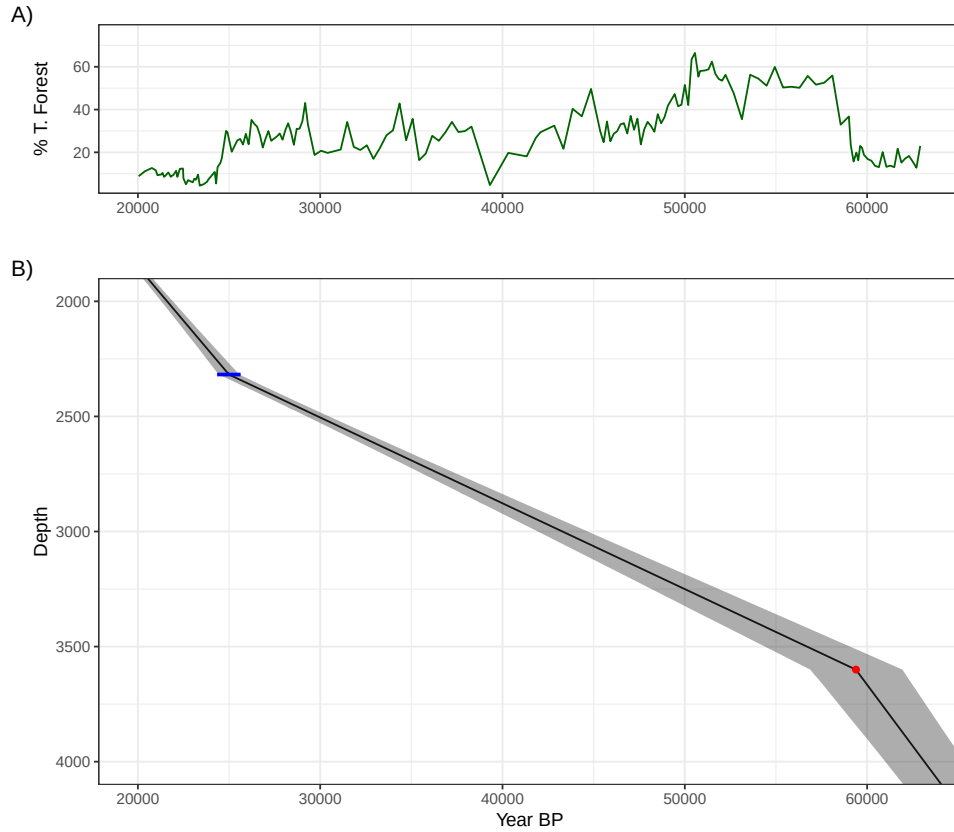


Figure 3.1: Example age model for the Ioannina I-284 pollen record, used for validation in Chapter 2. Panel A shows % Temperate Forest pollen for the Ioannina core (green line). Panel B shows the existing age model for Ioannina I-284 between 20 and 65 kyr BP (Sánchez Goñi et al., 2017). The existing chronology is constructed by linear interpolation between a single radiocarbon date (blue line) around 25 ky calBP and a tie point (red dot) at 59 kyr BP, resulting in an unrealistic linear age model for MIS 3. As a result, the pollen signal and its timing are uncertain, and may not be appropriate as validation for downscaled models in Chapter 2.

depth corresponding to 59 kyr BP (Sánchez Goñi et al., 2017). The result of this process is a completely linear age model for MIS 3 (Figure 3.1). Given some of the challenges apparent in dating these records (Jones et al., 2013; Staff et al., 2019; Tzedakis et al., 2002; Wulf et al., 2004, 2018), a quantitative tuning approach to this and other pollen record’s age models might be useful.

Dynamic time warping (DTW) and DTW-like algorithms have been applied to a number of quantitative tuning and proxy synchronisation problems (Bay et al., 2010; Burstyn et al., 2021; Hagen & Harper, 2023; Hay et al., 2019; Lin et al., 2014; Lisiecki & Lisiecki, 2002; Sirocko et al., 2021). DTW is a signal processing algorithm originally designed for speech recognition, developed to time-normalise

speech recordings of the same words spoken at different speeds (Sakoe & Chiba, 1978). Broadly, DTW remaps the indices of one time series to those of another by considering all possible realignments, compressing or expanding parts of the signal by way of minimising its overall distance to the other (Giorgino, 2009; J. Li et al., 2016). At present, no comprehensive applications of DTW to terrestrial palaeoenvironmental signals over MIS 3 are published, nor are any experiments assimilating GCM-derived predictions and observed palaeoenvironmental data by DTW.

In this chapter, three high-resolution European pollen records covering MIS 3 are aligned by DTW to downscaled predictions at their respective locations, as produced in Chapter 2. DTW is applied with an adjusted “step pattern” known as Minimal Variance Matching (MVM), having convenient constraints for quantitative tuning in palaeoenvironmental science (Latecki et al., 2007). An overview of the Dynamic Time Warping algorithm is given in Section 3.2.1, followed by a description of the MVM step pattern and its application to palaeoenvironmental synchronisation and the records used in this study. Results are presented in Section 3.3, where the realigned records’ age models are compared to their existing chronologies. Discussion of these comparisons and recommendations for future studies are given in Section 3.4.

3.2 Materials and Methods

3.2.1 DTW Algorithm

A brief overview of DTW is provided here. For detailed descriptions of the DTW algorithm and more complex implementations see [Sakoe & Chiba \(1978\)](#), [J. Li et al. \(2016\)](#), [Giorgino \(2009\)](#), [Petitjean et al. \(2011\)](#), [Sardá-Espinosa \(2019\)](#), and [Burstyn et al. \(2021\)](#).

At the core of the DTW method, in its simplest format, is a comparison between two vectors: a *query* of length N , $X = [x_1, \dots, x_N]$; and a *reference* of length M , $Y = [y_1, \dots, y_M]$. To compare and align X to Y by DTW, an N -by- M distance matrix D is first constructed, with each element ($D(n, m)$) representing some relationship (f) between X_n and Y_m :

$$D(n, m) = f(X_n, Y_m) \quad (3.1)$$

Most commonly in DTW f is given by the Euclidean distance between X_n and Y_m : $D(n, m) = \sqrt{(X_n - Y_m)^2}$. However, as noted by [Giorgino \(2009\)](#), any suitable function can be used to define f , and therefore build the distance matrix D .¹

A “cumulative cost matrix” (C) is then built, using the distance matrix D . Again, various methods exist for defining C , but in DTW’s most common format, C is calculated at the indices n and m using adjacent values of D , at $n-1$ and $m-1$:

$$C(n, m) = D(n, m) + \min \left\{ \begin{array}{l} C(n-1, m) \\ C(n, m-1) \\ C(n-1, m-1) \end{array} \right\} \quad (3.2)$$

A path across C is then drawn, found by a least-cost algorithm, originating at $C(N, M)$ and working in reverse through the matrix to $C(1, 1)$. The indices of this least-cost path across C represent the *warping function*, that realigns the indices of X to Y (and vice-versa). The value at $C(N, M)$ can also be used as a

¹In Chapter 4, an adjusted version of DTW defines the function f using adjacent values of n and m , for example.

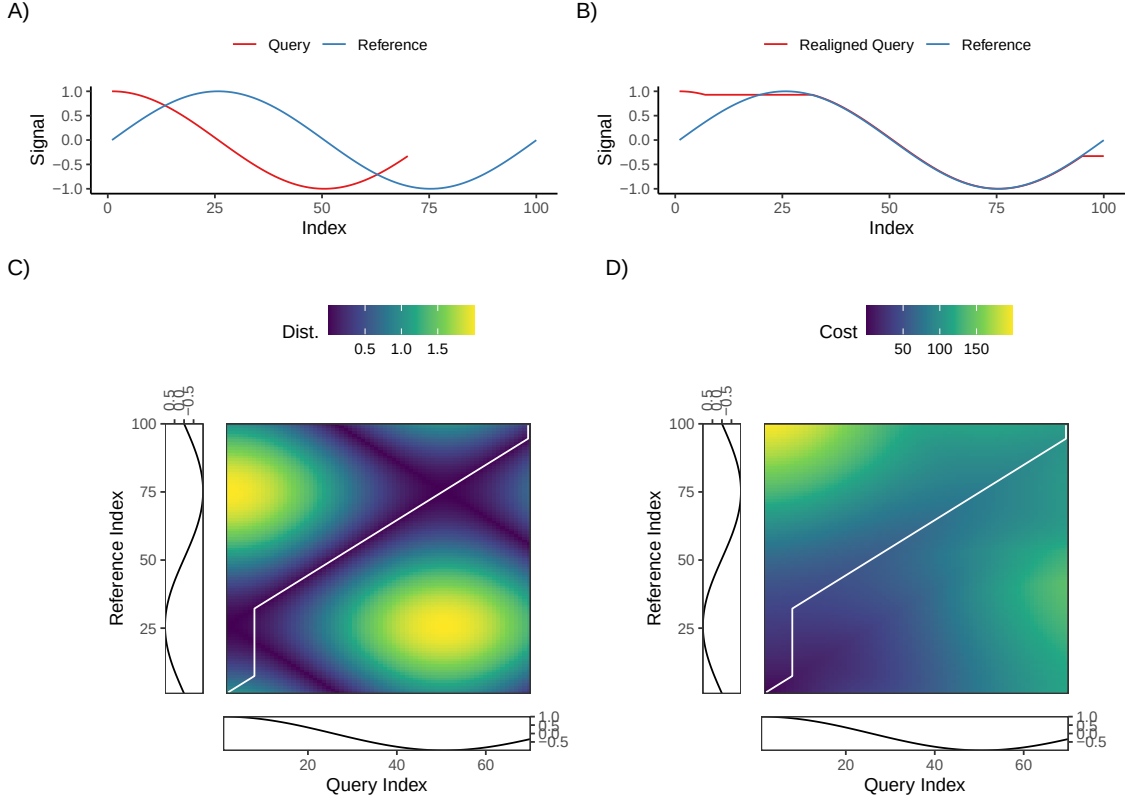


Figure 3.2: An example alignment problem and solution by Symmetric DTW. Panel A shows two signals: a reference sine wave (blue line); and a query (red line), to be aligned to the reference. Panel B shows the query realigned after Symmetric DTW. Panel C shows the distance matrix D (Equation (3.1)), here given by the Euclidean distance between the query and reference at all possible offsets. Panel D shows the cumulative cost matrix C constructed by the Symmetric step pattern (Equation (3.2)). The warping path, drawn by a least cost path across C and representing the realignment of the query indices to those of the reference, is overlain on Panels C and D in white.

measure of similarity between X and Y , as it represents the minimum cumulative cost of their alignment. In the above scheme, normalising the value at $C(N, M)$ is possible by dividing by $N + M$. Normalisation allows for comparison between alignments of different pairs of signals, which might have varying lengths. These measures are known as DTW distance ($C(N, M)$) and normalised DTW distance ($\frac{C(N, M)}{N + M}$).² An example of the application of Symmetric DTW to align two sine waves of varying length is shown in Figure 3.2.

DTW as implemented using the above scheme controls the shape of the warping path by the minimising function in Equation (3.2), while ensuring the following

²Normalisation is not always advisable, given that alignment rules defined by step pattern can restrict the shape of the warping path (Giorgino, 2009).

constraints (Folgado et al., 2018):

- *Monotonicity*, meaning that the warping path cannot have a negative gradient ($C(1, 1) \leq C(1, 2) \leq C(1, M)$, and $C(1, 1) \leq C(2, 1) \leq C(N, 1)$).
- *Boundaries*, meaning that the “heads” (X_1 and Y_1) and “tails” (X_N and Y_M) of X and Y are aligned to each other, with realignment only possible between these points.
- *A step size condition*, meaning that elements of X and Y cannot be omitted in alignment. This also means that each step in the warping path is always horizontal, vertical, or has a gradient of 1 (Figure 3.2).

It is possible to adjust Equation 3.2 and remove the step size condition above, by expanding or constraining the candidate elements of C . For example, indices can be adjusted to include values at $n - 1$ and $m - 2$:

$$C(n, m) = D(n, m) + \min \left\{ \begin{array}{l} C(n - 1, m) \\ C(n - 1, m - 1) \\ C(n - 1, m - 2) \end{array} \right\} \quad (3.3)$$

By allowing for a greater number of candidate elements of C , DTW can be expanded to allow large numbers of possible matches between indices of X and Y , allowing for different gradients in the warping path. Or, a cost can be applied to certain indices, making certain gradients of warping path more or less likely. For example, by doubling the value at $C(n - 1, m - 1)$, linear alignment is made less likely:

$$C(n, m) = D(n, m) + \min \left\{ \begin{array}{l} C(n - 1, m) \\ C(n, m - 1) \\ C(n - 1, m - 1) \cdot 2 \end{array} \right\} \quad (3.4)$$

By varying parameters such as those in Equations 3.3 and 3.4, an alignment can be intentionally biased toward compression or expansion of X to Y . Effectively, these adjustments influence and control the gradient of the warping curve (Figure 3.2). The DTW package in R allows control over these parameters by a scheme

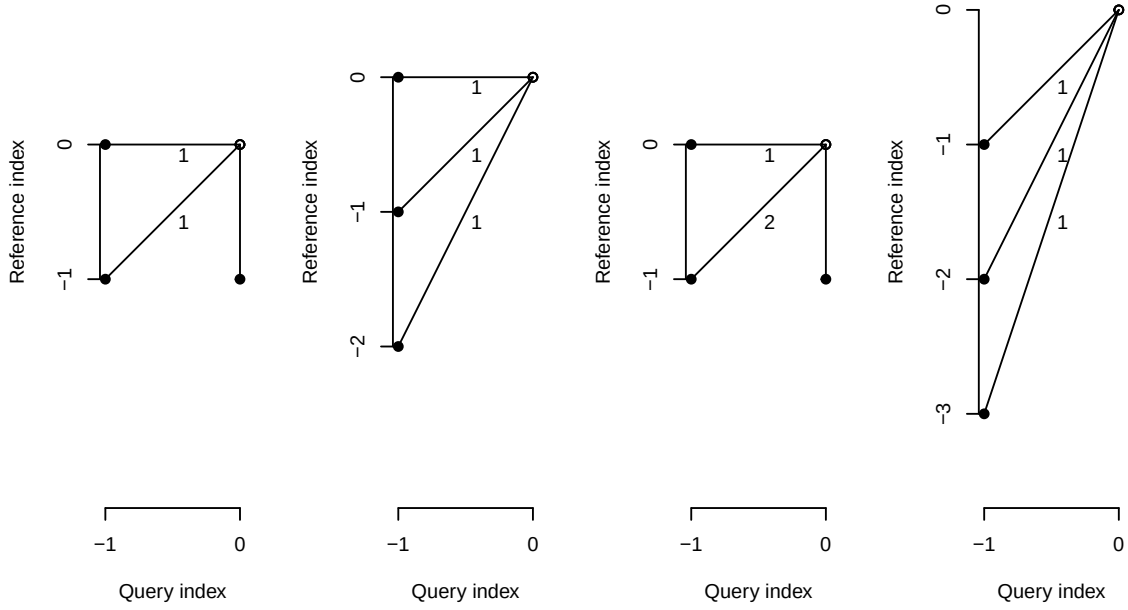


Figure 3.3: Comparison of step patterns in DTW: Symmetric (left, Equation (3.2)); Asymmetric (middle left, Equation (3.3)); Symmetric 2 (middle right, Equation (3.4)); MVM with elasticity (L) = 3 (right, Equation (3.5)). Black dots represent candidates for alignment to black circles. Numbers next to lines connecting dots and circles are multipliers for weighting alignment.

known as *step patterns* (Giorgino, 2009). A number of pre-defined step patterns exist, as well as further constraints on alignment not relevant to palaeoenvironmental synchronisation (Sakoe & Chiba, 1978). Equations 3.2, 3.3 and 3.4 describe the well-known *Symmetric*, *Asymmetric* and *Symmetric 2* step patterns, respectively (Figure 3.3; Giorgino, 2009).

3.2.2 DTW for Palaeoenvironmental Synchronisation and the MVM Algorithm

Having considered the most basic version of the DTW method, we will now consider some of the adaptations for applying DTW to palaeoenvironmental synchronisation.

Under equation 3.2, X and Y are both arbitrarily compressible and expandable in warping, with no constraint on how many indices of X can be aligned to a single index of Y , and vice-versa (see Figure 3.2). For the case of using DTW for synchronisation of palaeoclimate proxies, this is inconvenient. For example, if

aligning a pollen record at ~ 300 year sampling resolution (X) to an ice core record at ~ 20 year sampling resolution (Y), it is reasonable to assume that no two (or more) elements of X can represent the same single measurement in Y (i.e., if X_n is aligned to Y_m , X_{n+1} cannot also be aligned to Y_m). This drawback of the DTW algorithm has been overlooked in previous applications of the method to proxy synchronisation (Burstyn et al., 2021; Hagen & Harper, 2023; Hay et al., 2019; Lin et al., 2014; Lisiecki & Lisiecki, 2002).

An extension to DTW was developed by Latecki et al. (2007) known as the Minimal Variance Matching (MVM) algorithm. MVM was designed for partial shape matching, in an attempt to account for DTW's sensitivity to outliers in pattern recognition applications. Its structure is also convenient for palaeoenvironmental synchronisation, in that it ensures each element of X is aligned to a unique element of Y .

As explained by Giorgino (2009), MVM can be thought of as DTW with a user specified and size-variable step pattern. The parameter L , known as *elasticity*, offers some control over the flexibility of the warping path, in defining how many elements of Y are considered to give each value of C (Latecki et al., 2007). Under MVM, the cumulative cost matrix C is defined as:

$$C(n, m) = D(n, m) + \min \left\{ \begin{array}{c} C(n, m-1) \\ C(n-1, m-1) \\ C(n-2, m-1) \\ \dots \\ C(n-L, m-1) \end{array} \right\} \quad (3.5)$$

where L is the maximum number of indices of Y that can be considered for alignment, implying the additional constraints $L \geq \frac{M}{N}$ and $M \geq N$. Where the lengths of X and Y are equal, DTW by MVM only has one possible solution, a linear alignment. An example alignment problem and solution by DTW with MVM is shown in Figure 3.4.

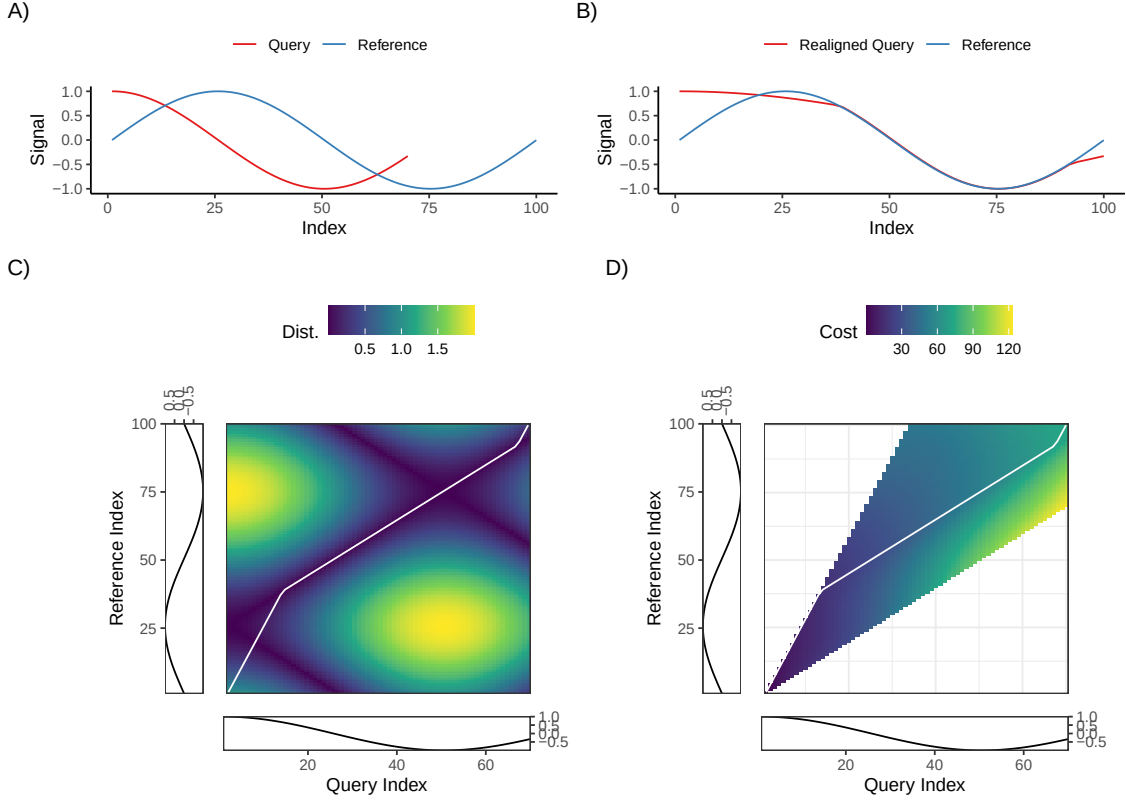


Figure 3.4: The example alignment problem shown in Figure 3.2 and its solution by DTW with the Minimal Variance Matching (MVM) step pattern. Panel A shows the query and reference signals. Panel B shows the realigned query. Panel C shows the distance matrix D , as for Figure 3.2. Panel D shows the cumulative cost matrix C constructed by the MVM step pattern with $L = 3$ (Equation (3.5)). In Panel C and D, the warping path is overlain in white. Under MVM, the warping path cannot be vertical or horizontal, meaning that no two indices of the query can be aligned to a single index of the reference, and vice-versa. Implicitly, areas of the cost matrix are undefined under MVM.

Where the length of Y is greater than that of X (i.e., $M > N$), portions of the reference vector must be skipped by the warping path, with the maximum size of these “jumps” controlled by the parameter L . Considering the alignment of two palaeoenvironmental signals by MVM, it is necessary to set L to a realistic value, to avoid improbable jumps in age models where L is too large. In this study, L was defined as $L = 2 \left\lfloor \frac{M}{N} \right\rfloor + 1$, where $\lfloor \cdot \rfloor$ denotes rounding to the nearest integer. This means that the window of re-alignment for a given point of X was limited to approximately twice the ratio of the resolution of X to the resolution of Y .

For example, if Y is a vector of securely dated measurements of length 600,

regularly spaced over a 30 kyr period, the sampling resolution of Y is 50 years. A query vector X of length 200, meanwhile, that perhaps is *not* securely dated, but is known to represent the same 30 kyr period, can be assumed to have a sampling resolution of 150 years. On average and over the whole 30 kyr period, there are three measurements of Y for every measurement of X . A linear solution would therefore align each value of X to every third value of Y . Aligning X to Y by DTW with MVM, L would be set to $2 \left\lfloor \frac{600}{200} \right\rfloor + 1 = 7$, meaning the maximum possible “jump” in the resulting warping path would be 7 indices of Y , or 350 years.

Using the principles of the DTW algorithm and MVM step pattern described above, univariate signals of three pollen records covering MIS 3 were aligned to temperature and precipitation predictions produced by the downscaling experiment in Chapter 2. In the following section, these records’ and their signals are described.

3.2.3 Palaeoenvironmental Records

Three terrestrial pollen records were aligned by DTW to the downscaled climate model predictions made in Chapter 2. All three records show variability in the proportions of arboreal and woody taxa among their pollen assemblages over MIS 3 (e.g., *Quercus*, *Abies*, *Fagus*). As shown in Chapter 2, these fluctuations have been condensed to a univariate signal of either percentage arboreal pollen or percentage temperate forest pollen, published alongside their chronologies in the ACER Pollen Database (Sánchez Goñi et al., 2017) and by Staff et al. (2019). A short description of these records and their existing age models is provided below.

For all three pollen records, the most recent published chronologies (Sánchez Goñi et al., 2017; Staff et al., 2019) provided the start (60 kyr BP) and end (30 kyr BP) of regions of the core considered for alignment. These approximations were assumed to be “correct” for the purposes of this study. Query (X) and reference (Y) signals were both scaled before alignment by DTW, by subtracting the mean of each respective vectors from each measurement and dividing by the vectors’ standard deviations.

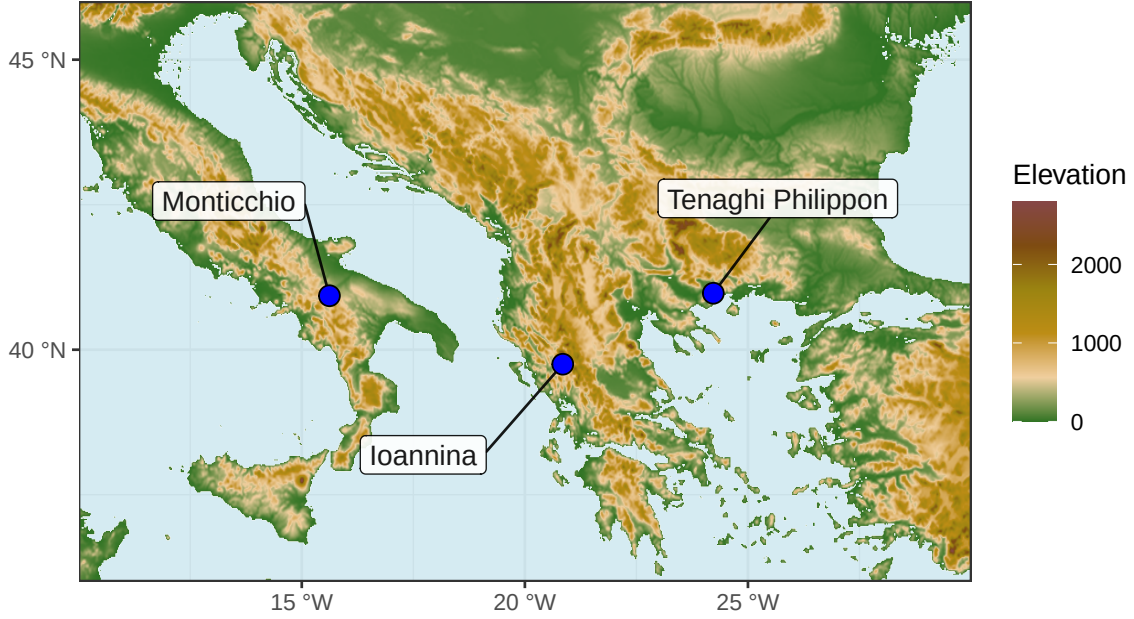


Figure 3.5: Locations of the pollen records used in this study.

3.2.3.1 Downscaled Reconstructions

Predictions for temperature and precipitation over MIS 3 made in Chapter 2 were extracted from the downscaled models at the locations of each pollen core, shown in Figure 3.5. For each pollen record, separate alignments were made to temperature and precipitation predictions, resulting in two DTW-based age models for each core: one based on downscaled temperature, and another based on downscaled precipitation.

100 year time steps for downscaled temperature and precipitation predictions were linearly interpolated to 50 year resolution before alignment, to ensure to satisfy the condition $M \geq N$. Interpolating the reference records also allowed for the possibility of high-resolution sampling regions for cores, as well as variability in sedimentation rates.

For temperature, the *minimum* temperature of downscaled predictions (i.e., mean temperature of the coldest month) at each location was used as a climate proxy for alignment. Minimum temperatures showed greater variability in downscaling results than mean or maximum temperatures (see Chapter 2), and are known to have more influence than summer or mean annual temperatures on

the ecological distribution of plants, in general (Harrison & Sanchez Goñi, 2010). For precipitation alignments, the total annual precipitation at each core's location was used.

3.2.3.2 Tenaghi Philippon

The first pollen record considered was the TP-2005 core from Tenaghi Philippon in Northern Greece. At Tenaghi Philippon, 200 m of peat and lake sediments capture around 1.3 million years of vegetation dynamics (Tzedakis et al., 2002). TP-2005 is a 60 m long core recovered in 2005, representing the past 310 kyrs of vegetation changes at Tenaghi Philippon, sampled for pollen analysis at between 2 and 5 cm spacings (Fletcher et al., 2010; Müller et al., 2011).

TP-2005 is a well-studied and high-resolution archive of vegetation change in the eastern Mediterranean, showing abrupt changes in vegetation dynamics throughout the last glacial, with a cyclicity that, in some parts, and perhaps more clearly than the other records considered here, resembles the DO pattern of abrupt warming followed by gradual cooling (Figure 3.6). For TP-2005, these vegetation changes are most clearly seen in the percentage arboreal pollen (% Arboreal), and this variable was used as a climate proxy for alignment (Figure 3.6). Furthermore, this variable is currently the only published and accessible pollen data from the TP-2005 core (Staff et al., 2019).

As well as more clearly reflecting DO cycles, the independent chronology of TP-2005 might be argued as the most robust of the three considered here. The MIS 3 section of TP-2005 have been extensively dated since by radiocarbon since the core's recovery in 2005. Müller et al. (2011) presented 20 AMS dates along with the pollen stratigraphy, suggesting a sample resolution of around 100 years per sample across the majority of MIS 2 and MIS 3 (before 50 ky calBP). Cape Riva (Y-2), Campe Flegrei (Y-3), and Campanian Ignimbrite (Y-5) tephra were also identified in TP-2005, constraining the chronology further. Staff et al. (2019) updated the chronology by way of 44 new AMS dates for the section below the Y-5 tephra (~ 40 - 50 ky calBP). Wulf et al. (2018) combined seven tephra layers, 60

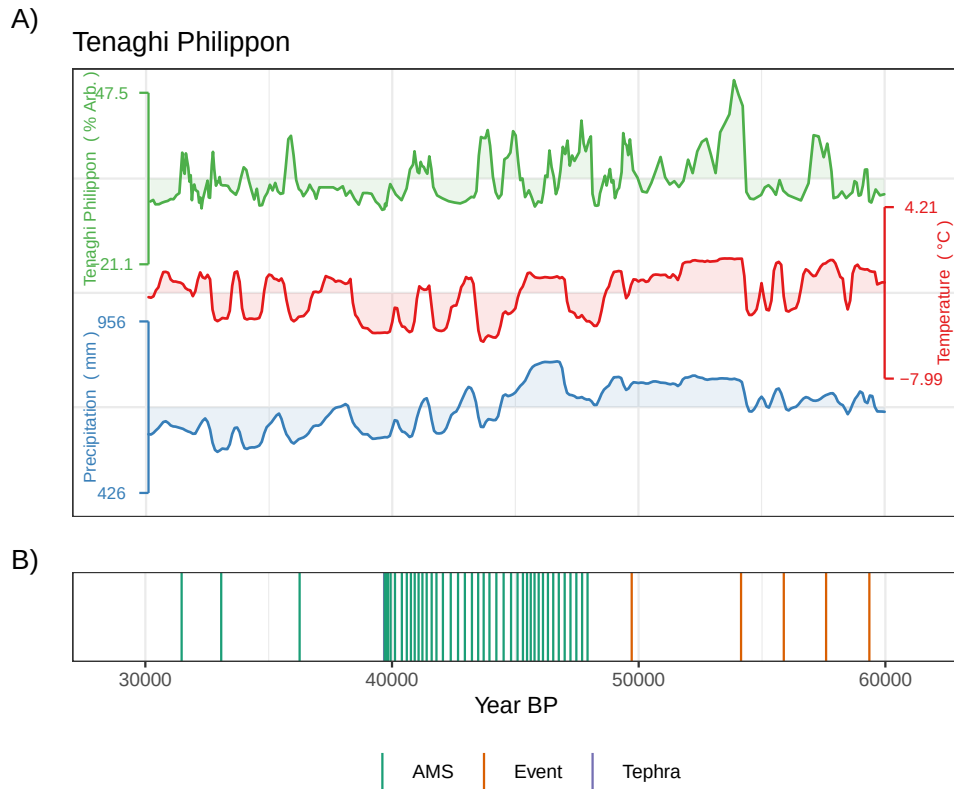


Figure 3.6: The Tenaghi Philippon pollen record and its existing chronology. Panel A shows % Arboreal pollen from the Tenaghi Philippon sediment core (green line), compared to downscaled temperature (red line) and precipitation (blue line) predictions. Vertical lines in Panel B shows locations of chronological information used to build the age model, by radiocarbon measurements (dark green), manual tuning ("Event" layers; orange), and tephra layers (purple).

of the AMS dates, and 20 palynological tuning points to build the chronology that underlies Figure 3.10.

3.2.3.3 Monticchio

One of the first long pollen records to be recognised as having patterns similar to Greenlandic DO events was from Lago Grande de Monticchio (Monticchio), a Maar lake formed in a volcanic crater in Southern Italy with sediments covering the last 132 kyrs (Allen et al., 1999; Fletcher et al., 2010). Pollen preserved at Monticchio provides a continuous record of vegetation changes for the last glacial period, measured at fairly high-resolution (~ 200 years per sample for MIS 3 and MIS 2; Allen et al., 1999). Large parts of the sequence are annually laminated (varved), and several tephra layers found in cores taken at Monticchio have been

dated by $^{40}\text{Ar}/^{39}\text{Ar}$ (Ramrath et al., 1999). Together, these constitute a chronology that is wholly independent to the Greenland records, though breaks in varved sections of the core mean some areas of the chronology hold some uncertainties (Allen et al., 1999).

The vegetation sequence at Monticchio reveals millennial-scale environmental changes for the area surrounding the lake. Proportions of woody taxa fluctuate dramatically throughout the sequence, with greater percentages of these taxa during warm periods, alternating to steppic vegetation during cold periods, being particularly pronounced changes are seen between 45 and 40 kyr BP based on the varve chronology (Figure 3.7). Around the time of Monticchio’s first coring and early study, parallel observations for the same period were made at marine sites of the Iberian margin (Fletcher et al., 2010; Sanchez Goñi & Harrison, 2010; Shackleton et al., 2000). Similar fluctuations at Monticchio were logically inferred to correspond to interstadial and stadial periods, as observed in marine cores and the Greenland chronology. For Monticchio, the percentage temperate forest species (% Temperate Forest) was the univariate signal used for alignment by DTW.

3.2.3.4 Ioannina

The pollen sequence of the I-284 core from Lake Ioannina in Northern Greece similarly covers the entire last glacial period and beyond, to around 250 ky BP (Tzedakis et al., 2002, 2004). Recent comprehensive tephra analysis was undertaken on a separate core (I-08) from Lake Ioannina for the last ~ 50 ky (McGuire et al., 2022), but unfortunately this has not yet been paralleled for the I-284 core. Dating the Ioannina cores by radiocarbon has also been difficult because of the “hardwater effect,” where old carbon of karstic origin introduces biases to radiocarbon measurements, as well as a general lack of appropriate material for dating for later parts of the sequence (e.g., macrofossils, macrocharcoal; Jones et al., 2013).

Due to these limitations, I-284’s most up-to-date chronology, as plotted in Figure 3.8, is based on tuning. This involved aligning the peaks of oscillations in its pollen record to those in $\delta^{18}\text{O}$ marine records from the Iberian margin, which

Monticchio

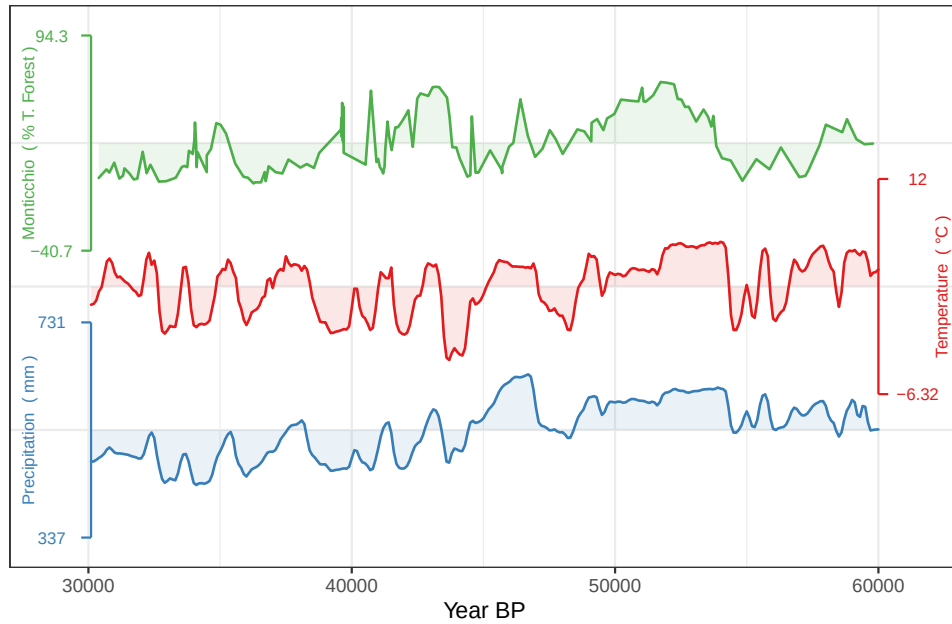


Figure 3.7: % Temperate Forest pollen from the Monticchio sediment core (green line) based on its existing age model, compared to downscaled temperature (red line) and precipitation (blue line) predictions. As the Monticchio age model is based on varve counting, no additional chronological information is shown.

in turn were tuned to the pattern of the Greenland record (known as a “stepping-stone” approach, [Fletcher et al., 2010](#)). As such, Ioannina presented an interesting opportunity to compare quantitative tuning by DTW against wholly manual tuning. Based on the tuned chronology, the sample resolution of pollen samples is estimated by [Fletcher et al. \(2010\)](#) to be 225 years per sample.

As for Monticchio, millennial-scale expansions and contractions of temperate vegetation seen at Ioannina have been linked to Greenland stadials and interstadials, though these shifts are less pronounced for MIS 3. At Ioannina, % Forest pollen was used for alignment to downscaled predictions.

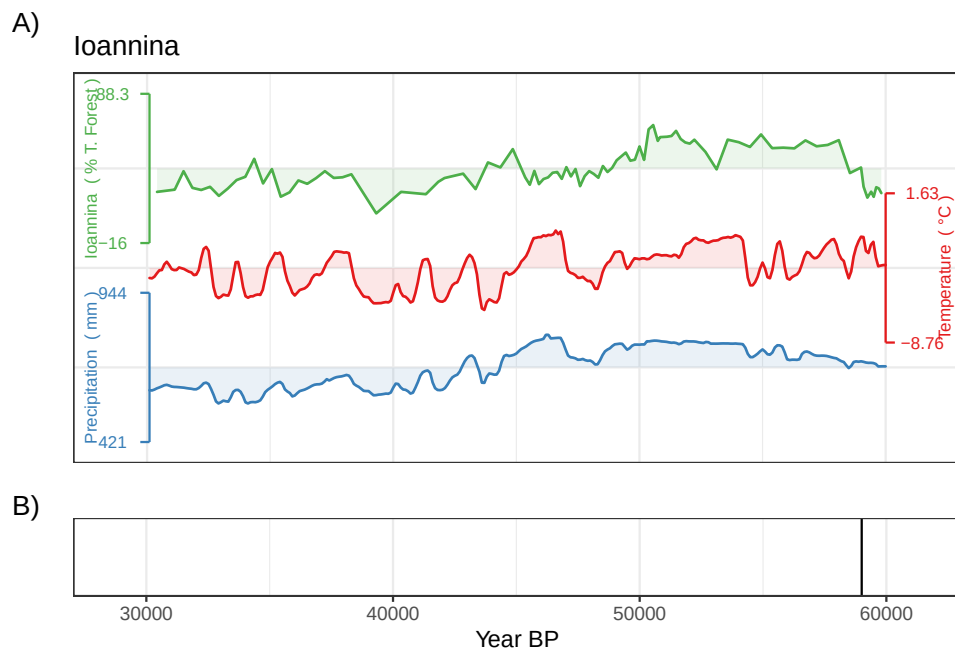


Figure 3.8: The Ioannina I-284 pollen record. Panel A shows % Temperate Forest pollen from the Ioannina sediment core (green line) based on its existing age model, compared to downscaled temperature (red line) and precipitation (blue line) predictions. Panel B shows the location of the single tuning point at 59 kyr BP used to build the Ioannina chronology.

Table 3.1: Results of alignment by DTW to downscaled temperature and precipitation predictions for the three pollen records considered. Column "Length" shows the length of the records between 60 and 30 kyr BP (MIS 3), and "Res" their resolution over the same period. L is the elasticity parameter used in the MVM step pattern. Columns "Dist" and "N Dist" refer to the DTW and normalised DTW distances in alignment, respectively.

Record	Variable	Length	Res.	L	Temperature		Precipitation	
					Dist	N Dist	Dist	N Dist
Tenaghi Philippon	% Arboreal	251	120	5	395	1.57	399	1.59
Monticchio	% Temperate Forest	153	196	9	203	1.33	202	1.32
Ioannina	% Temperate Forest	110	273	11	146	1.33	155	1.41

3.3 Results

A summary of results for alignment by DTW to downscaled temperature and precipitation for the three pollen records considered in this study is presented in Table 3.1. DTW distances were higher for higher-resolution records, but after normalising by query length (N) no clear relationship was found between the resolution of the pollen records and their normalised DTW scores.

The lowest DTW distance of all alignments, or, the highest similarity between queries and references after realignment, was found for alignment between % Temperate Forest pollen and downscaled precipitation at Monticchio (normalised DTW distance = 1.32). Some differences between DTW scores for temperature and precipitation were apparent at each location, most notably at Ioannina. At Tenaghi Philippon and Ioannina, alignment to downscaled temperature reconstructions gave a lower DTW score than precipitation, reflecting greater similarity (or, lower cumulative distance) between pollen signals and downscaled temperature.

Original and DTW-based age models are shown in Figure 3.9. None of the DTW age models appear unphysical, but separation from the original age models was apparent in all cases. There were no unrealistic jumps in sedimentation rate in any of the DTW-based age models, suggesting that the choice of $L = 2 \left\lfloor \frac{M}{N} \right\rfloor + 1$ in the MVM step pattern was appropriate. Alignment by downscaled temperature

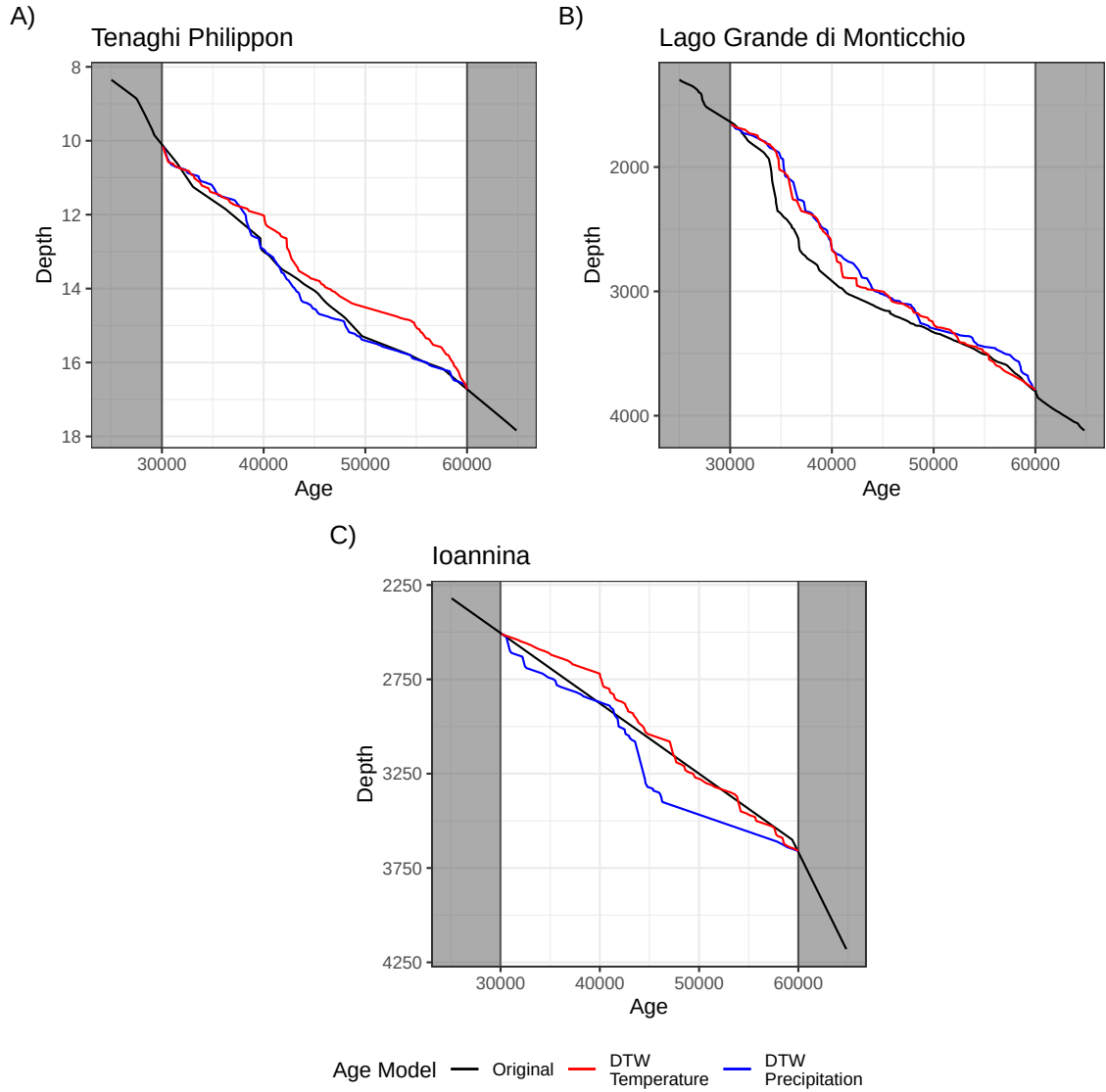


Figure 3.9: Comparison between original age models (black lines) and age models produced by DTW alignment to downscaled temperature (red lines) and downscaled precipitation (blue lines), at Tenaghi Philippon (Panel A), Monticchio (Panel B), and Ioannina (Panel C). Grey shading before 60 kyr BP and after 30 kyr BP show regions not considered for alignment.

and Tenaghi Philippon and alignment by precipitation at Ioannina showed the greatest differences to the original age models.

3.3.1 Tenaghi Philippon

The DTW age model aligned by temperature at Tenaghi Philippon showed greatest differences at the start of MIS 3 (Figure 3.9A), where alignment by DTW suggested the original age model underestimates the age of the core. At 15 m depth, this

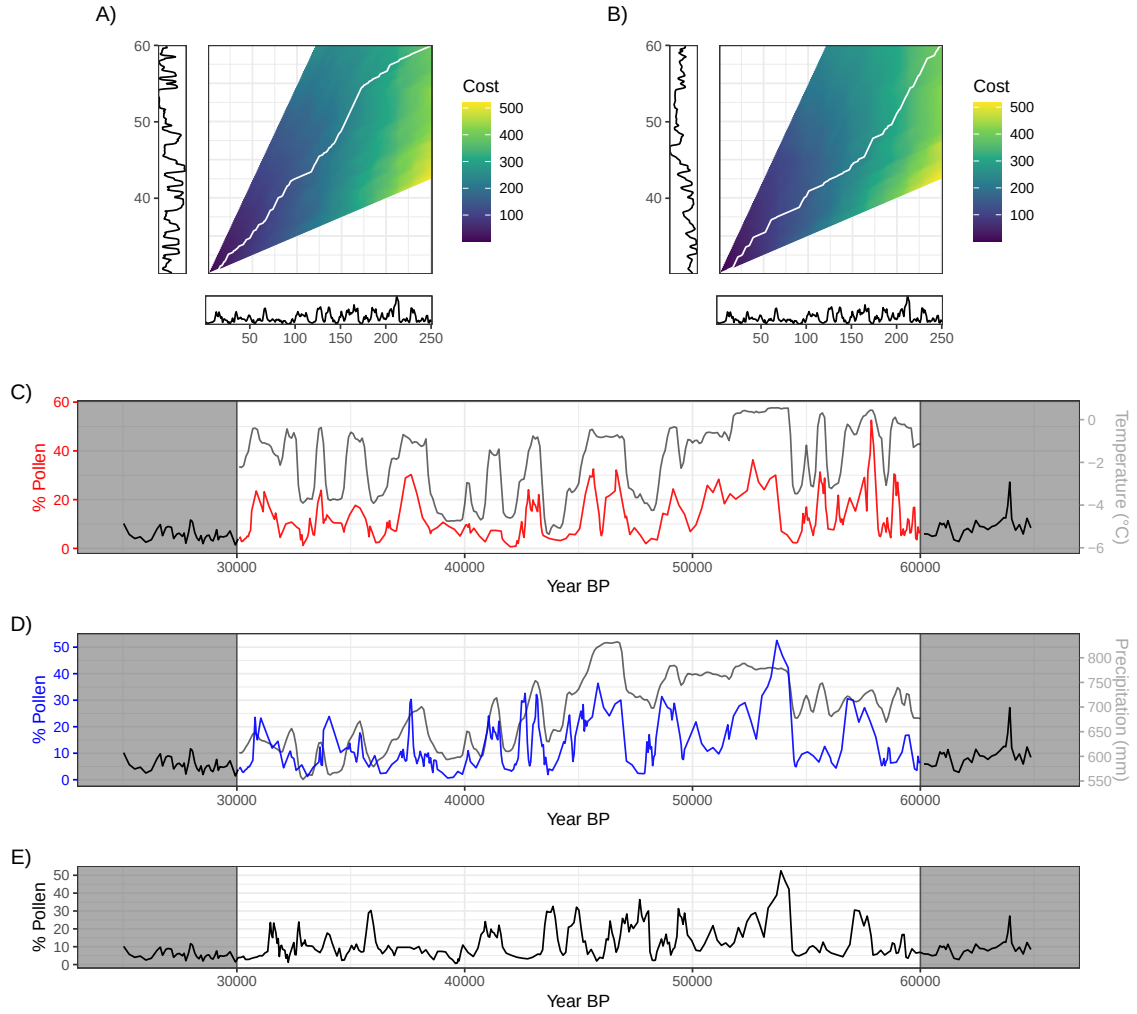


Figure 3.10: Realigned signals for Tenaghi Philippon. Panels A and B shows the cumulative cost matrix (C) with warping path for alignment by DTW to downscaled temperature and precipitation, respectively. Panels C and D show the realignment of the Tenaghi Philippon pollen record (red and blue lines, respectively, left y-axes) to the downscaled temperature and precipitation predictions (grey lines, right y-axes). Panel E shows the pollen signal plotted to its original chronology.

difference was around 7 kyr. The warping path for temperature (Figure 3.10A) showed compression during this period, with a relaxed gradient for these indices of the warping path. The effect of this compression was a “squeezing” of the original pollen signal to indices corresponding to 60 and 55 kyr BP (Figure 3.10C). Expansion in the signal was shown for later periods, but some compression was also apparent at ~ 43 kyr BP.

The DTW model aligned by precipitation showed much closer similarities to the original Tenaghi Philippon age model, particularly for the earlier part of MIS

3. A clear peak in arboreal pollen at the onset of GI 14 (~ 54 kyr BP) was aligned with an increase in precipitation predicted by the downscaling model (3.10D), as tuned in the chronology by Wulf et al. (2018). Later parts of MIS 3 were more compressed under the precipitation model, with relaxed gradient in the warping path after 40 kyr BP. Downscaled precipitation showed similar patterns to the % Arboreal pollen at Tenaghi Philippon after ~ 48 kyr BP after alignment (Figure 3.10D). A decline in precipitation over this period predicted by the downscaling model and HadCM3B, punctuated by DO oscillations, reflected a similar trend in Tenaghi Philippon's % Arboreal pollen. The DTW precipitation age model for the same period followed the original model, which in turn suggested an agreement with the radiocarbon chronology at Tenaghi Philippon (Staff et al., 2019).

3.3.2 Monticchio

At Monticchio, both precipitation- and temperature-based DTW alignments suggested the original varve-based age model consistently underestimates the age of the core through MIS 3. The maximum offset between DTW-based and original age models was around 3 kyr at 25 m depth (Figure 3.9B). Earlier parts of MIS 3 were more closely aligned to the original chronology for both DTW models.

Alignment by DTW to downscaled temperature for Monticchio showed both compression and expansion throughout the core, with compression particularly evident in the middle of MIS 3 (~ 45 kyr BP, Figure 3.11A). This period aligns with GI 12, GS 12, and GI 11, where the pollen signal was particularly “warped,” resulting in a variable gradient in the age model. Although this age model was not unrealistic, alternations between compression and expansion in DTW-alignment over this period imply variability in sedimentation rate, and in this in turn requires some sedimentological explanation.

The alignment of the % Temperate pollen signal to downscaled precipitation by DTW showed a much clearer relationship between realigned signals (Figure 3.11D). As for Tenaghi Philippon, an overall decreasing trend in precipitation fitted remarkably well to a similar pattern in the Monticchio pollen signal. The

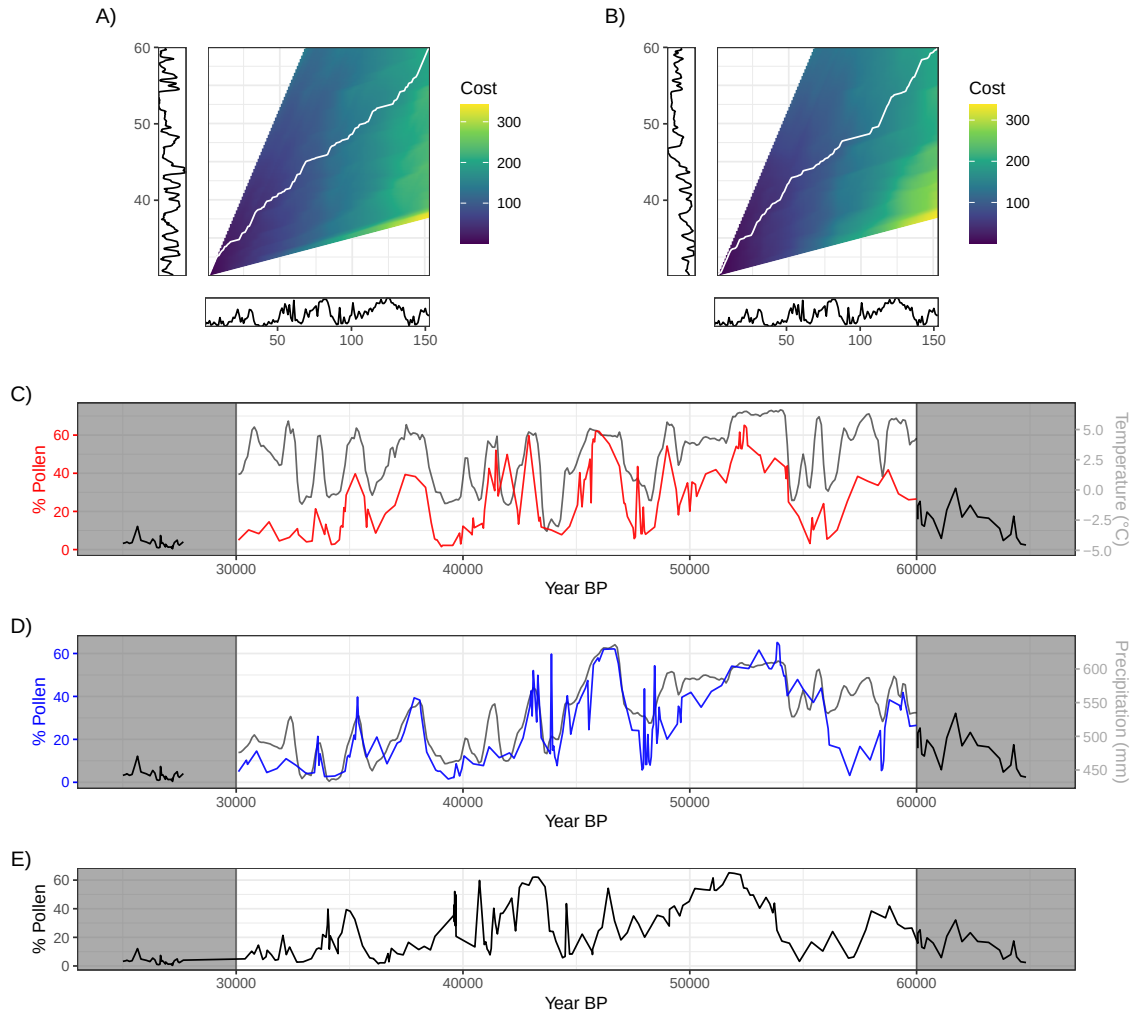


Figure 3.11: Realigned signals for Monticchio. Panels A and B shows the cumulative cost matrix (C) with warping path for alignment by DTW to downscaled temperature and precipitation, respectively. Panels C and D show the realignment of the Tenaghi Philippon pollen record (red and blue lines, respectively, left y-axes) to the downscaled temperature and precipitation predictions (grey lines, right y-axes). Panel E shows the pollen signal plotted to its original chronology.

shape of the warping path, as for alignment to downscaled temperature, was variable, with periods of compression and periods of expansion (Figure 3.11B). GI 12 and GI 11 are particularly well-represented after alignment, but GI 10 appears to be “missing” in the Monticchio signal, according to alignment by DTW to downscaled precipitation.

3.3.3 Ioannina

At Ioannina, the two age models produced by alignment of the pollen signal to downscaled temperature and downscaled precipitation by DTW were very different (Figure 3.9C). The temperature-based age model showed greater similarity to the original, particularly from the start of MIS 3 until around 45 kyr BP. Alignment to downscaled precipitation suggested the original age model overestimates the age of the core for all periods of MIS 3 except around 40 kyr BP. At around 45 kyr BP, an offset of around 5 kyr was shown between the original age model and that produced by alignment to downscaled precipitation.

The temperature-based alignment by DTW showed similarities to the original age model at the start of MIS 3, with some compression of the pollen signal between 60 and 50 kyr BP (Figure 3.12A). None of the GI or GS periods were particularly well-pronounced in the pollen signal adjusted by downscaled temperature-based DTW, other than a gradual decrease in % Temperate Forest pollen over GI 14 and GI 13 (Figure 3.12B). GIs 10 to 5 (between ~ 40 and 30 kyr BP) were poorly represented in the signal aligned by downscaled temperature.

The precipitation-based alignment showed a considerable expansion for the early part of MIS 3 (60 to 48 kyr BP), with the shape of the warping path at this period being as steep as the MVM algorithm allows. This would imply that the sampling resolution of the Ioannina core during this period is around 550 years, twice that suggested by Fletcher et al. (2010) and Tzedakis et al. (2002). Following this period, between ~ 47 and 41 kyr BP, considerable compression was shown, with the overall decline in % Temperate Forest pollen aligned to a steep decline in precipitation (Figure 3.12D). During this period, the corresponding age model suggested a much slower sedimentation rate (Figure 3.9C). The precipitation warping path for these two periods (60 to 48 kyr and 47 to 41 kyr BP) showed the most dramatic changes between compression and expansion in comparison to other cores.

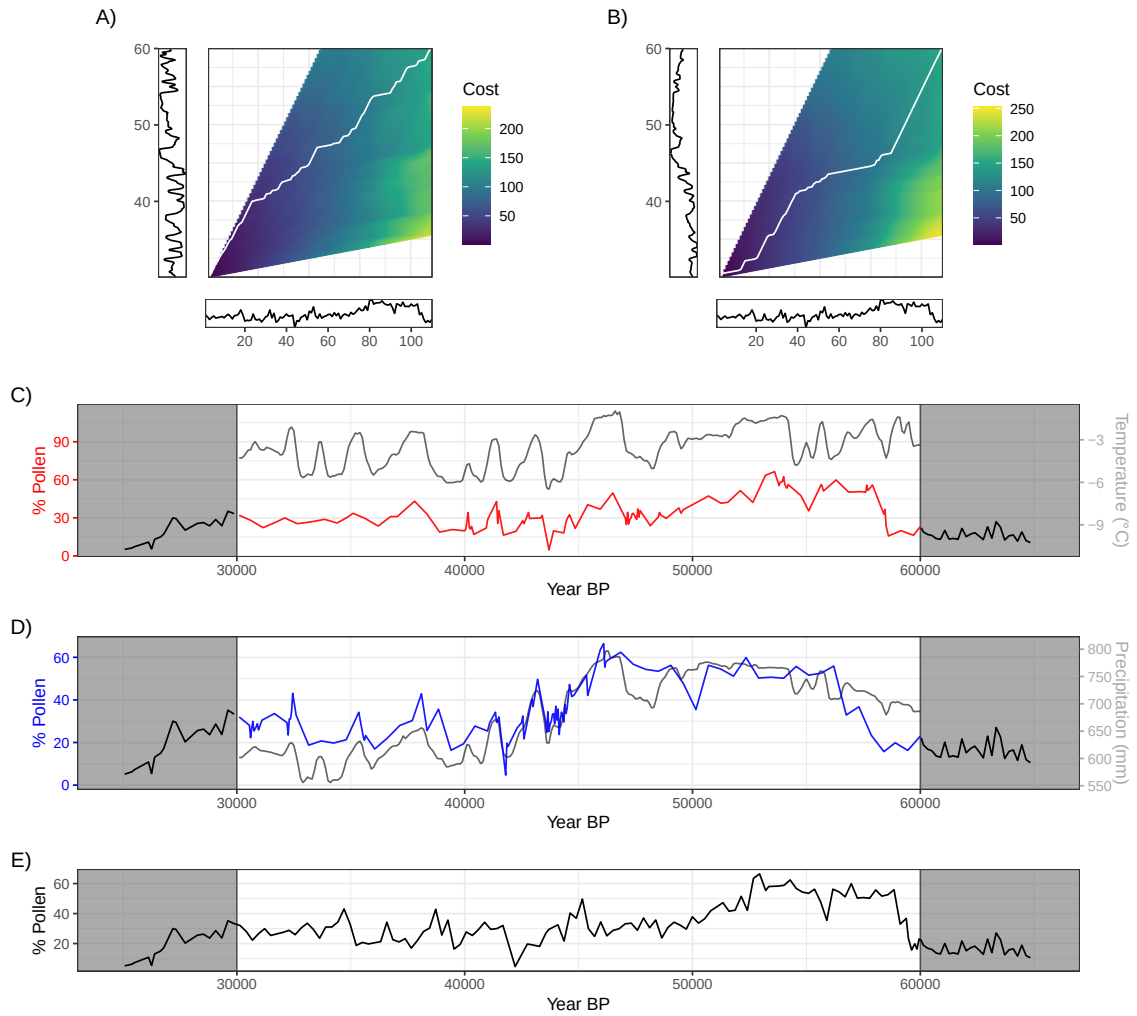


Figure 3.12: Realigned signals for Ioannina. Panels A and B shows the cumulative cost matrix (C) with warping path for alignment by DTW to downscaled temperature and precipitation, respectively. Panels C and D show the realignment of the Tenaghi Philippon pollen record (red and blue lines, respectively, left y-axes) to the downscaled temperature and precipitation predictions (grey lines, right y-axes). Panel E shows the pollen signal plotted to its original chronology.

3.4 Discussion

This chapter presents an experiment in quantitative palaeoenvironmental synchronisation using the DTW algorithm with an MVM step pattern (Giorgino, 2009; Latecki et al., 2007). Proxies for palaeoenvironmental variability from three pollen cores from southern and south eastern Europe were synchronised to downscaled temperature and precipitation predictions made in Chapter 2, producing alternative age models for cores at each location (Allen et al., 2000; Fletcher et al., 2010; Tzedakis et al., 2002).

Alignment by DTW improved the relationship between pollen signals and the downscaled climate variables, the latter reflecting the pattern of variability captured in Greenland ice core records (Kindler et al., 2014; Rasmussen et al., 2014). Plausible age models were produced by DTW in all cases. The study confirms the potential of signal processing algorithms as an alternative to manual tuning strategies, while demonstrating a possible method to assimilate modelled data and palaeoenvironmental observations. However, some key assumptions underlie the approach taken, while there are areas for improvement to the DTW algorithm and its application to proxy synchronisation for MIS 3.

Comparisons between existing age models and those produced by DTW alignment showed significant differences over various parts of MIS 3. As well as these differences, DTW alignments to temperature and precipitation predictions produced distinct age models, with different DTW scores (Figure 3.9, Table 3.1). While both minimum monthly temperature and total annual precipitation are known to have significant influence on plant species distribution (Harrison & Sanchez Goñi, 2010), further investigation into which signal is more appropriate for alignment is necessary before ascertaining which alignment is more relevant.

The existing age models for the three records considered were produced by a variety of methods: at Tenaghi Philippon, by a combination of radiocarbon dates, tephra layers, and manual tuning strategies (Wulf et al., 2018); at Monticchio, by the core’s varve chronology (Allen et al., 1999; Fletcher et al., 2010); and at

Ioannina, by linear interpolation between a radiocarbon date ~ 25 kyr calBP and a tie point at 59 kyr BP, the latter tuned to the MD95-2042 marine core (Shackleton et al., 2000; Tzedakis et al., 2002). Reasons for discrepancies between existing and realigned age models might be illuminated by direct comparisons between DTW-aligned age models and the data that contributed to the existing age models (e.g., radiocarbon dates, tephra layers and their dates), as well as dates at these cores that did *not* contribute to existing age models (e.g., radiocarbon measurements at Monticchio). As for the previous chapter, assessing the reasons why each individual record’s existing chronology does not agree with its respective pollen signal is beyond the scope of this study. However, assuming the pollen signals *should* align with those of other records, possible reasons for these discrepancies include: reservoir effects or contamination in radiocarbon dates at Tenaghi Philippon, as highlighted by revised dates by Staff et al. (2019); counting errors or missing varves propagated through the chronology at Monticchio; and poorly constrained linear interpolation between tuning points at Ioannina.

If implementing DTW for proxy tuning, age estimates for the start and end of the query should be well-constrained. Here, existing age models for each record were “sliced” at 30 and 60 kyr BP, with the period in between considered for alignment. A significant limitation of this study is the assumption that the existing age models are “correct” at these points. This is of course problematic and somewhat circular given that the motivation for realignment by DTW was uncertainty in existing age models. This constraint was somewhat unavoidable, necessitated by the chronological limits of the downscaling experiment of Chapter 2, where GCM simulations were downscaled between 30 and 60 kyr BP. As the DO-adjusted HadCM3B surfaces by Armstrong et al. (2019) extend to 0 kyr BP it would be possible to downscale the entire time series between 60 and 0 kyr BP, and realign larger portions of the pollen records to downscaled predictions. However, under the above scheme, it would still be necessary to “chop” the pollen records at 60 kyr BP, or arbitrarily choose (i.e, manually tune) a point of the pollen signal believed to represent 60 kyr BP.

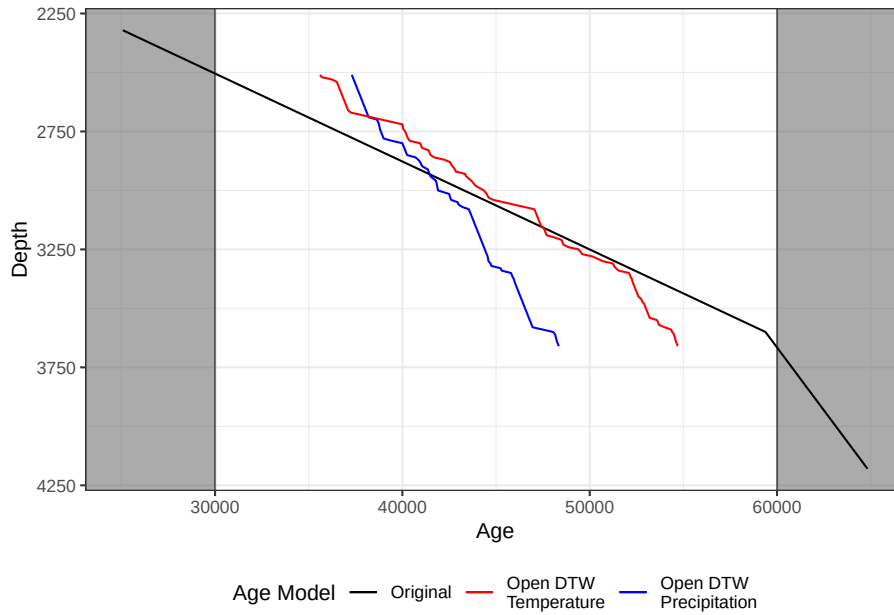


Figure 3.13: An example of DTW alignment for Ioannina where parameters "open.end" and "open.begin" are used. When the resolution of the query is relatively low, and without constraining the start and end of the query to those of the reference, the query signal is likely to be compressed to subsections of the reference.

An alternative approach would be to make use of the `open.end` and `open.begin` arguments of the DTW package (Giorgino, 2009). Effectively, this applies DTW iteratively by a sliding window, finding the query's optimum alignment *within* the reference signal. The beginning and end of the query vector can be aligned to any points of the reference, relieving the constraint of fixed alignment at the "heads" and "tails" of signals (Section 3.2.1). However, in experimental applications of open-end and open-begin DTW, the resulting age models are unrealistic: compression is likely when the resolution of the query is relatively low in comparison to the reference, as shown for Ioannina in Figure 3.13.

A key challenge for the application of DTW to proxy synchronisation is that the method as implemented in this chapter does not account for any uncertainties attached to palaeoenvironmental records. Such uncertainty might be chronological (see Figure 3.1), or be associated with the palaeoenvironmental signal itself (e.g., mass spectrometer instrument error in $\delta^{18}\text{O}$ records). The NGRIP record used to correct the HadCM3B simulations holds a chronological error of around 2 kyr at the beginning of MIS 3, for example, while radiocarbon measurements can also bring

considerable chronological uncertainty. Given that the DTW method as applied above is deterministic, these uncertainties are overlooked. Other studies have successfully incorporated uncertainty into DTW alignment, at much broader chronological scales. [Lin et al. \(2014\)](#) implement the “Match” algorithm, a DTW-like technique for stratigraphic alignment ([Lisiecki & Lisiecki, 2002](#)), using probabilistic methods, constrained by estimates for sedimentation rates that are employed as priors. Using a hidden Markov model, “HMM-Match” results an alignment solution as for DTW, as well as an estimate of its error. Future applications of DTW for proxy alignment for MIS 3 should similarly focus on methods - probabilistic or otherwise - to account for uncertainty in palaeoenvironmental records. Incorporating existing chronological information, uncertainty, and outlier analysis to the alignment process might also be possible (e.g., [Figure 3.7](#) and [Figure 3.6](#)).

A broader question is whether pollen records for MIS 3 are appropriate in their resolution for quantitative tuning. While DTW has been applied to various other palaeoenvironmental records, such as Greenlandic ice core records or $\delta^{18}\text{O}$ measurements of marine sediments ([Lin et al., 2014](#); [Lisiecki & Lisiecki, 2002](#)), these applications are at much broader chronological scales, capturing and aligning millennial variability, rather than centennial or decadal. [Burstyn et al. \(2021\)](#) used hierarchical DTW to cluster and align sub-annual signals in speleothem records, but here the application was at much higher resolution with greater sampling frequency than proxies for MIS 3 (~ 6000 measurements over approximately 18 years of speleothem growth). The sampling resolutions of the Tenaghi Philippon, Monticchio, and Ioannina cores are relatively high for the Late Pleistocene, but when focusing on MIS 3 their resolutions may be insufficient for comparison and alignment. Another related assumption is that any peak of a given signal represents the true local maximum for the proxy, but this might be erroneous due to the cores’ relatively low sampling resolution. Again, employing probabilistic methods to account for uncertainty in depth, for example, might alleviate this problem for MIS 3 ([Lougheed & Obrochta, 2019](#); [Muschitiello et al., 2020](#); [Muschitiello & Aquino-lopez, 2023](#)).

A still broader and perhaps more significant question is whether records can be expected to show synchronicity over MIS 3. Some debate surrounds whether tuning is appropriate for Late Pleistocene records - quantitative, manual, or otherwise - due to its inherent disregard of any “leads and lags” between large-scale atmospheric and oceanic shifts, their evolution in the earth system, and the localised ecological responses to these changes (Austin & Hibbert, 2012; Blaauw, 2012; Rasmussen et al., 2014; Tzedakis et al., 2004). A recent compilation of speleothem records has suggested that for mid-latitudes of the northern hemisphere, abrupt shifts over MIS 3 can assumed to be synchronous (Corrick et al., 2020), but some local discrepancies remain (Adolphi et al., 2018). Radiocarbon dating is particularly challenging for marine records (Lougheed et al., 2013), and this coupled with relatively large uncertainties for ^{14}C dates over MIS 3, means that robustly testing proxy synchronicity over MIS 3 with precision is immensely challenging.

To some extent, tuning to downscaled predictions or simulations of climate by a GCM can offer some alleviation to this problem, offering some independence from and alternative perspective on the assumption of synchronicity, through their numerical representation of the earth climate system. Downscaled predictions for MIS 3 oscillations show variability in the magnitude of their changes at different locations, confirming that the GCM-based predictions do simulate local variability and regional climate responses to large-scale changes (see Chapter 2). However, understanding their evolution and time-dependent questions relating to leads and lags in the climate system is not possible by the Armstrong et al. (2019) data. The HadGCM3B simulations were produced by a “snapshot” approach to climate modelling, rather than being a continuous simulation, while Armstrong et al. (2019) further corrected the simulations using the NGRIP record and time scale. In other words, using the downscaled simulations from Chapter 2 to realign pollen records is more or less the same approach as tuning to the NGRIP record, and suffers from the same drawbacks and assumptions of synchronicity as other tuning strategies.

Alongside the more theoretical challenges relating to palaeoenvironmental proxy synchronisation, other areas for improvement to DTW for quantitative tuning are

more technical. First, both pollen and downscaled temperature and precipitation signals were scaled before their alignment by DTW. Scaling allows for the magnitude of signals to be compared more readily, but approaches using transfer functions to reconstruct a climate signal from pollen percentages are also available (Braak & Juggins, 1993; Chevalier et al., 2021; Salonen et al., 2019). Using a transfer function to translate a pollen assemblage to minimum or winter temperatures, for example, might also mean that model-data comparisons and validation in Chapter 2 are more meaningful: if age uncertainties can be accounted for by DTW, then absolute differences between modelled and observed reconstructions could be quantified. However, pollen transfer functions are by no means perfect: they are rarely validated adequately and carry significant uncertainty (Chevalier et al., 2020). The danger of building “models upon models” is that the uncertainties and errors of each modelling stage in the process are propagated to the next, becoming increasingly difficult to locate and quantify.

Second, the approach taken in this study does not account for variation in the sampling resolution across the depth of a core. If, for example, a portion of a pollen core is sampled at very high frequency, it is likely that this section of the chronology will be “stretched” by DTW, with an anomalously higher deposition rate represented by the resulting age model. A potential remedy for this would be varying L in Equation (3.5) by sample depth, to account for changes in sampling resolution.

DTW has clearest potential in proxy synchronisation for cores where no other independent age estimates are available over MIS 3, such as at Ioannina. For pollen records with existing radiocarbon dates, tephra layers, or varve chronologies, DTW might allow some assessment of the *difference* between their signal and age model throughout the core, rather than replacing their age models altogether. Combining this approach between signals could be particularly useful for assessing the chronology and spatial variability of leads and lags in the climate system, for example.

3.5 Conclusion

This chapter presented an experiment in the quantitative alignment of palaeoenvironmental records to downscaled predictions made in Chapter 2. The signal processing algorithm DTW was applied to three relatively high-resolution pollen signals from Europe, using the MVM step pattern. Alignment by DTW produced realistic age models for all three records, but revealed discrepancies between the pollen signals’ original chronologies and those produced by DTW. Furthermore, aligning to temperature and precipitation predictions produced distinct age models in each case. Discerning which signal is more relevant in relation to the % Temperate Forest or % Arboreal pollen signals requires further investigation, or the use of pollen transfer functions to translate signals to climate variables.

Discrepancies between existing and realigned age models highlights the challenge of building robust and accurate age models for palaeoenvironmental records over MIS 3. Rather than the DTW-based age models “updating” the existing ones, they should be used to further understand the *distance* between radiocarbon, varve-based, or tephrochronological estimates and the pollen signals. If peaks in pollen signals can be assumed to have been synchronous with DO events, alignment by DTW highlights areas for chronological improvement. If, on the other hand, the existing chronologies are considered “correct,” alignment by DTW might be one way to quantify the effect of “leads and lags” in the climate system, and the ecological responses to abrupt climate shifts over MIS 3.

The poor resolution of records over MIS 3 was also highlighted as a drawback of using DTW for proxy alignment. In the following chapter, a high-resolution application of DTW in palaeoenvironmental science is considered, surrounding the problem of cross dating in dendrochronology.

4

Dendrochronological Applications of Dynamic Time Warping

Abstract

Dendrochronology is a widely used dating method in archaeology and palaeoenvironmental science. Dating by dendrochronology involves aligning a sample of annually resolved tree ring width measurements to a “master sequence,” an averaged chronology of ring widths for a given region and species. A number of methods exist for finding the closest index match in a master sequence, the most common using moving-window Pearson Correlation and *t*-Test statistics (Baillie & Pilcher, 1973; Wigley et al., 1987). However, in some circumstances local disturbances in growing conditions or counting errors during tree ring measurement can mean that sequences contain multiple or missing rings for a given year. In such cases, alignment by correlation can be problematic and an absolute date difficult to obtain (Schweingruber, 1988).

In Chapter 3, the Dynamic Time Warping (DTW) algorithm was applied to palaeoenvironmental records in order to realign their chronologies and avoid biases of manually tuning records. In this chapter, an adjusted version of the DTW algorithm is applied to master sequence matching, providing an alternative method

for dendrochronological cross-dating for cases where additional or missing rings are present. A correlation-based distance measure and moving-window approach are incorporated in alignment by DTW. The algorithm also inherently provides a similarity metric between the two signals, which is used to create an adjusted t -Test statistic, allowing comparison to alignment by Pearson Correlation.

The performance of the adjusted DTW algorithm is assessed using a well-defined chronology of kauri tree rings from northern New Zealand ([Palmer et al., 2016](#)). Random sub-sequences of varying length are sampled from the kauri series to simulate examples of added and missing rings in cross-dating and compare performances between the two methods of alignment ($N = 18,000$). Comparisons show that when no missing rings are present, cross-dating by DTW gives similar results to cross-dating by Pearson Correlation, and that DTW consistently outperforms Pearson Correlation in accuracy where rings are added or removed. Results show that DTW has excellent potential in dendrochronology, particularly for cases where no strong match is found between a sample and master sequence by correlation.

4.1 Introduction

Dendrochronology is the study of tree ring widths and their variation over time for dating purposes (Schweingruber, 1988). Due to accelerated cell growth during the growing season and slower growth at other times of the year, many species of tree from mid-latitudes show visible rings in cross section. Annual changes in environmental conditions manifest in changes in the widths of growth rings, as well as their characteristics, such as isotopic composition or density (Loader et al., 2019; McCarroll & Loader, 2004; Pilcher et al., 1984; Schweingruber, 1988). Examples of environmental variability affecting tree ring growth are increases or decreases in seasonal temperatures or precipitation rates, changes in soil quality, or changes in light availability within the canopy. Tree rings grown under stressed environmental conditions are generally narrower than those grown under optimal conditions.

Distinctive patterns can be identified across tree ring sequences from multiple trees grown in the same region and under the same conditions. By combining these regional tree ring sequences, long time series of variability in tree ring growth can be created, and in turn used as proxy records for past climatic variability (Cook et al., 1994). These series can be extended back in time from modern trees, extrapolated from known-age samples, or “floating” tree ring chronologies can be built for collections of related sequences that lack absolute dates. Very long species-specific chronologies of tree ring variability have been constructed for many mid-latitude regions (Díaz et al., 2001; Ferguson, 1969; Friedrich et al., 2004; Panyushkina et al., 2010; Zhang & Qiu, 2007), and these can be averaged into long archives known as master sequences (Figure 4.2). Master sequences provide valuable records of past environmental variability, and their chronologies are typically very well-constrained due to their annual resolution and low uncertainties (Cook et al., 1995; García-González, 2008).

Master sequences are also useful for dating tree ring samples of unknown age. Using a master sequence for a given region, species, and time period, a tree ring sample can be dated by alignment, by searching for the region where the

sample and master sequence “agree” most. This technique is known as cross-dating (Baillie & Pilcher, 1973; Wigley et al., 1984), with the last ring in the sequence giving a *terminus post quem* for the date of the tree’s felling or death. An advantage of this method for archaeology and palaeoenvironmental science is that, in comparison to other dating methods such as radiocarbon dating, very precise and absolute dates with no or very little uncertainty can be obtained for samples, providing the master sequence is similarly well-constrained (Aitken, 1990). For this reason, tree ring analysis and dendrochronology have been fundamental for the development of radiocarbon dating methods, providing independent and high-resolution chronologies from which to calibrate radiocarbon measurements (Pearson et al., 2022; Plicht et al., 2020).

A number of methods have been developed for cross-dating since the method’s formalisation in the 1920s, from visual alignment, through graphical methods such as “skeleton plotting,” to more complex numerical and probabilistic methods (Pearson et al., 2022). Today, cross-dating is typically performed numerically, and involves searching for locations where some measure of similarity between the sequence and master sequence the two is maximised. This approach can be carried out using a number of software packages (e.g., Bunn, 2008; Cook & Holmes, 1996; Grissino-Mayer, 2001; Holmes, 1983) and is generally not excessively computationally expensive. Common measures of similarity used in dendrochronology are the Pearson correlation coefficient, the Student’s *t*-Test, and the *Gleichläufigkeitskoeffizient* (GLK; Schweingruber, 1988).

One of the main challenges in cross-dating a sample is the presence of missing or added rings (D. Li, 2022). Where environmental conditions are exceptionally poor, a tree may fail to grow an annual ring at all, resulting in a “missing” ring for that year. There are many environmental reasons for missing rings for a given year, such as: excessively high or low temperatures during the growing season; seasonal droughts; volcanic eruptions; or excessive pollution and smoke due to industrial activity (Grundmann et al., 2008). “False” rings in a sequence are also possible, where two rings are grown in one year. False rings are most frequently

induced by a drought during the growing season, interrupting the ring-growing process and causing a reversion to winter-like cell growth patterns. If and when the drought ceases later in the growing season, large cell growth returns and the tree appears to have grown two rings in one year ([Copenheaver et al., 2006](#); [Dyer & Fritts, 1978](#)). In addition to environmental variability, counting errors or duplicate measurements during data collection can mean that tree ring datasets contain missing or added rings, due to human error.

Both missing and false rings can be problematic in cross-dating, as alignment by Pearson correlation, t -Test, or GLK all assume that the time series of ring measurements in a tree ring sample represent annual increments, that can be sequentially paired to a subsequence of the master sequence of the same length ([Baillie & Pilcher, 1973](#); [Bunn, 2008](#); [Wigley et al., 1987](#)). In their simplest format, the most common similarity measures for cross-dating do not readily account for missing or additional rings.

To successfully align samples containing missing or false rings to a master sequence using the above methods, qualitative examination by a dendrochronologist is required to identify suspect areas of the sequence and adjust the sequence alignment accordingly. Alternatively, incremental calculations of a distance measure can be calculated across the sample and master sequence to identify locations that are likely to contain a missing or false ring ([Bunn, 2010](#)). Again, this requires some qualitative assessment of incremental comparisons before an adjusted alignment can be reached ([Grissino-Mayer, 2001](#)). Analysis of missing and false rings in the popular dendrochronology program COFECHA is possible by comparing segments of user-specified length to the master sequence ([Grissino-Mayer, 2001](#)). COFECHA then flags segments containing potential missing or false rings, indicating that significantly higher correlation scores are found at adjusted indices. However, to identify the location of a missing or false ring, users must repeatedly run the program using shorter segment lengths, and pinpoint the exact index by visual or graphical inspection.

As outlined in Chapter 3, Dynamic Time Warping (DTW) is a signal processing algorithm that compresses or stretches one signal time-ordered signal to the optimum alignment to another. DTW has clear potential for cross-dating in dendrochronology, particularly for cases where missing or false rings are present in a sample. Key advantages of DTW for dendrochronology are that: the two signals can vary in length, resolution, speed, or magnitude; a distance metric is inherently calculated by the algorithm, allowing comparison between multiple alignments; and the algorithm is relatively simple and intuitive to understand. The potential for using DTW in dendrochronology was recognised soon after the method’s development (Kruskal, 1983), and since then several studies have presented DTW-like algorithms for cross-dating in dendrochronology. Wenk (2003) uses an *edit distance* metric, which operates in a similar method to DTW, developing on Van Deusen (1990)’s edit distance. However, the sample size in Wenk (2003)’s study is small, limiting its conclusions. Other applications of DTW in dendrochronology include Zheng et al. (2021)’s application of DTW to cluster dendrochronological time series from the Hindu Kush Himalaya, and Oriani et al. (2022)’s use of DTW in tree ring detection and counting. As yet, no comprehensive experiments have been published that explore the application of DTW (or, DTW-like) algorithms in for cross-dating in dendrochronology.

This study presents an alternative method for cross-dating in dendrochronology using an adjusted version of the DTW algorithm. Section 4.2 describes the dataset used and design of the experiment, summarising the Pearson correlation approach to cross-dating and introducing the changes made to the DTW algorithm. Section 4.3 presents compared results produced by the experiment, focusing on the accuracy and any biases apparent either method’s performance. Finally, Section 4.4 and Section 4.5 summarise and discuss the key findings from this study and suggest improvements to the DTW algorithm for its future application to cross-dating in dendrochronology.

4.2 Materials and Methods

4.2.1 Towai Farm Dataset

The data set used for this study was taken from Palmer et al. (2016) and is available from the International Tree-Ring Databank (<https://www.ncei.noaa.gov/access/paleo-search/study/20750>). Palmer et al. (2016) present a data set of 92 tree ring sequences taken from 39 sub-fossil kauri (*Agathis australis*) logs discovered at Towai Farm, New Zealand (35° 30' S, 174° 10' E, Figure 4.1). From stratigraphic evidence, the “cohort” of kauri trees were deposited within a similar time period, but did not die at the same time.

The Towai series and its chronology were built by measuring ring widths across radial strips from cross sections cut from the sub-fossil logs, power-transforming the individual sequences, detrending each one by 200-year spline subtraction, then standardising sequences by the “signal free” method (Melvin & Briffa, 2008). cross-dating by Pearson correlation was then performed on the detrended and standardised sequences using the program COFECHA (Grissino-Mayer, 2001). Of the 39 logs recovered at Towai Farm, 37 were successfully cross-dated by correlation to create a floating chronology, presented in Figure 4.2.

The Towai Farm tree ring series is exceptionally well correlated, with an average inter-series correlation of 0.67, and only 9 sequences needed to surpass a recommended expressed population signal (EPS) of 0.85 (Palmer et al., 2016; Wigley et al., 1984). The whole series spans a 1451-year period, and a 1010-year period with an EPS > 0.85. The Towai series has been extensively dated by radiocarbon, with the 1010-year period dated to 12959-11949 calBP (Hogg et al., 2016). A significant ring width downturn is identified between ~12500 and 12380 calBP, within the period of rapid cooling known as Greenland Stadial 1 (GS-1) over Greenland, or the Younger Dryas over Europe Palmer et al. (2016).

Given its relatively long length and high EPS value, the Towai series is an ideal data set to compare alignment methods and test performance in cases of added and missing rings. The floating chronology published by Palmer et al. (2016) is

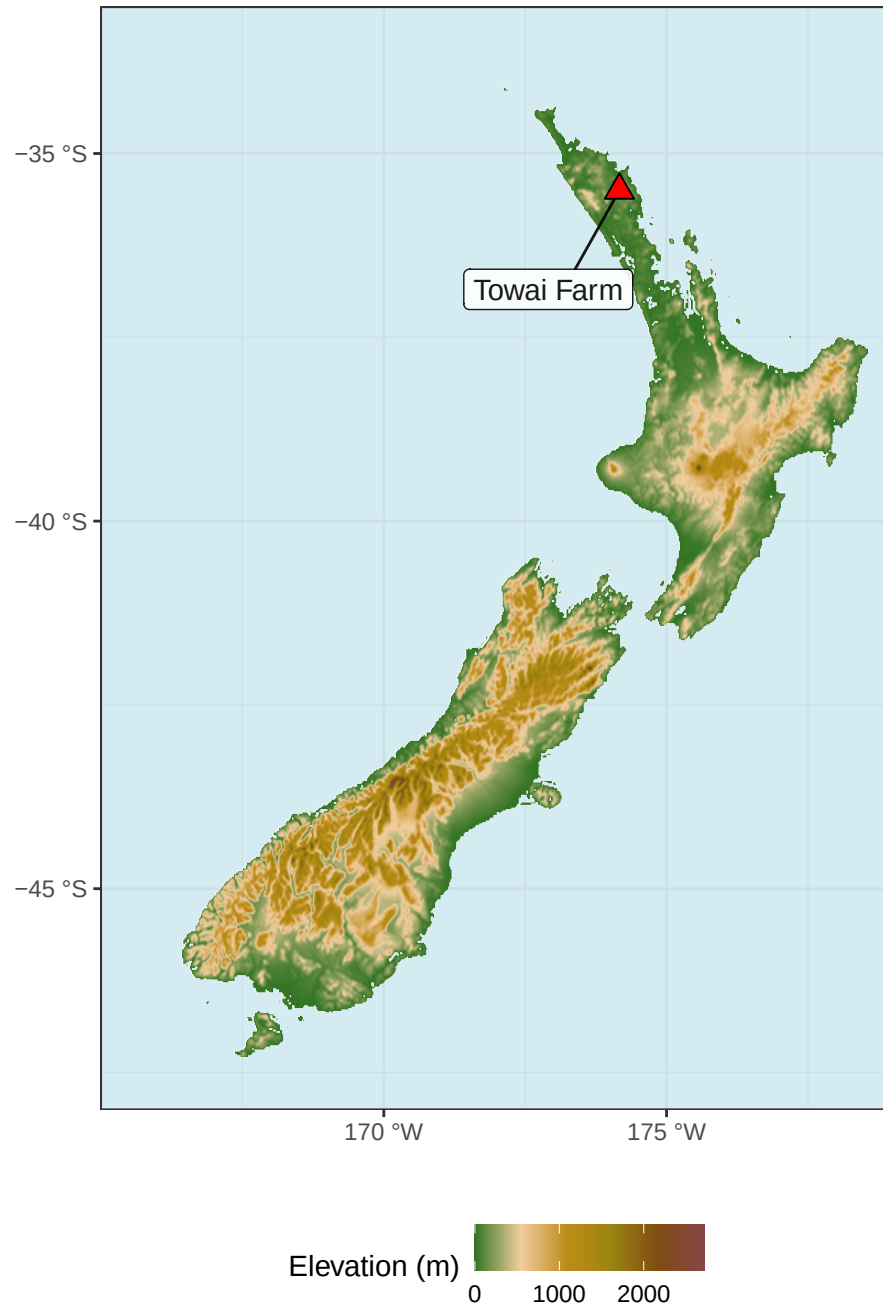


Figure 4.1: Location of Towai Farm (red triangle) in New Zealand.

used to create samples and master sequences to compare DTW and correlation cross-dating methods.

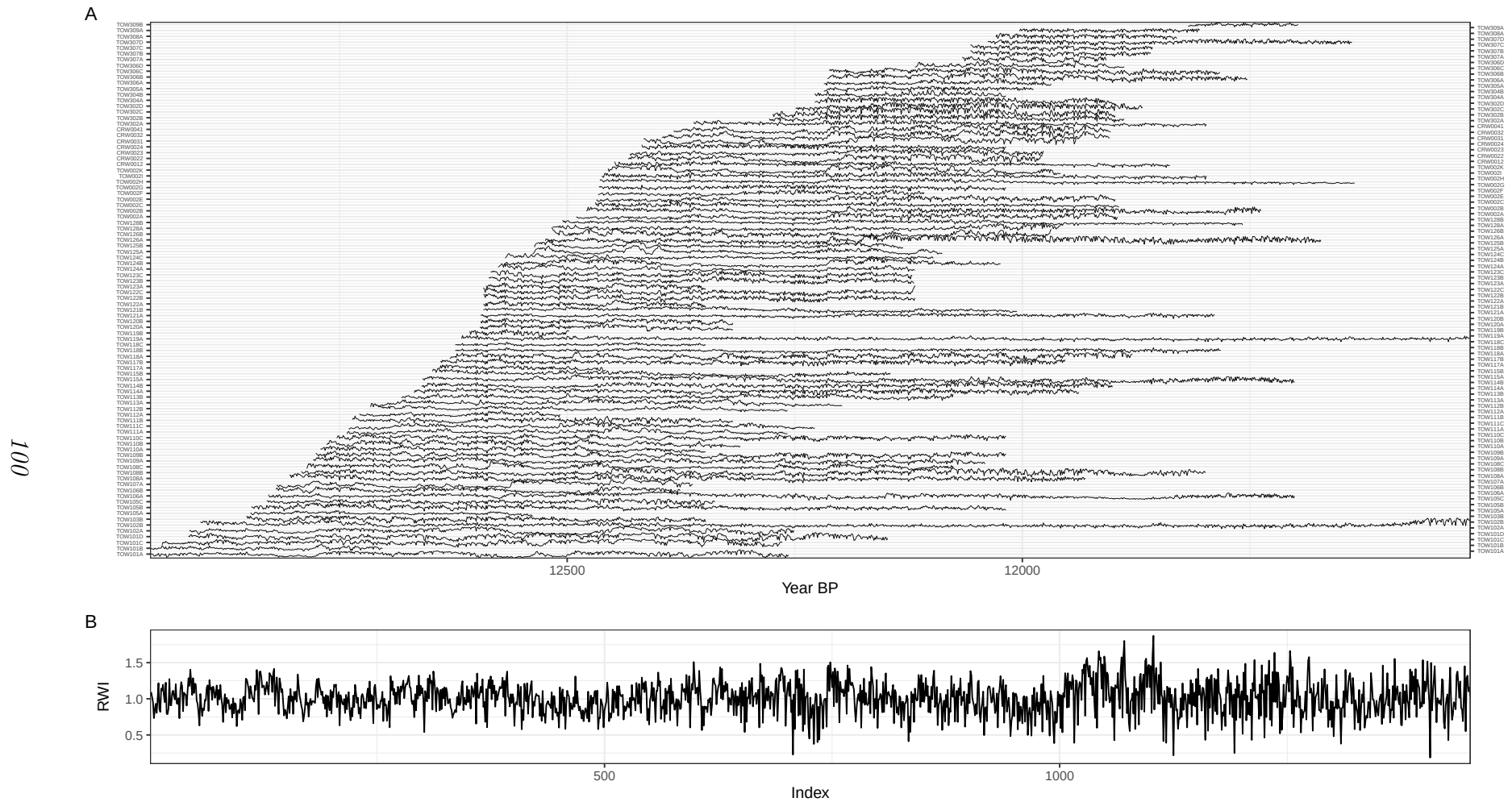


Figure 4.2: Towai Farm tree ring series and resulting master sequence (Palmer et. al., 2016). Panel A shows the individual sequences from the Towai series in their original alignment, from which samples of varying length were extracted and cross dated. Panel B shows the resulting Towai master sequence ring width indices (RWI), produced by log transformation, spline detrending and averaging across sequences (frequency response = 50%; wavelength = 50-years).

Table 4.1: Parameters and associated values used in this study.

Variable	Values	Notation	Unit
Sample		X	
Master sequence		Y	
Y offset		Δ	
Length of sample		n	
Length of Y_Δ		m	
Sequence lengths	20, 40, 60, 80, 100	l	Years
Number of rings added/removed	0, 1, 2, 3, 4, 5		Years
Spline frequency response	0.2		
Spline wavelength (master sequence)	30		Years
Spline wavelength (sample)	$\frac{n}{2}$		Years
DTW window size	0.1	k	
DTW non-linear penalty	0.7	z	
Accuracy threshold	10		Years

4.2.2 Sampling, Detrending, and Missing and Added Rings

Table 4.1 gives a summary of parameters used to generate samples for cross-dating comparisons between DTW and correlation.

To create each sample for cross-dating, samples were randomly extracted from the Towai series, with rings added or removed to mimic samples that are typically difficult to date by correlation. First, 0.1 mm was to all measurements to accommodate cases where ring widths were zero, before log-transforming the entire series. A tree ring sequence from the log-transformed Towai series (e.g., TOW002H) was randomly selected, before randomly selecting a continuous subsequence from the selected sequence, of length l . To explore the performance of cross-dating methods for varying sequence lengths, l was varied between 20 and 100 (Table 4.1).

Numbers of removed and added rings were varied between 0 and 5. Removed rings were randomly selected from sample indices. Added rings were generated by

randomly generating ring widths between minimum and maximum values within the sample, then inserting these values at random indices. For cases with removed and added rings, rings were removed before rings were added.

Before cross-dating, the log-transformed ring width measurements for each sample were detrended by spline filtering. Detrending is applied in dendrochronology to remove non-climatic signals from ring width measurements, as well as standardise measurements to make comparisons between trees meaningful (Wigley et al., 1987). A tree generally grows faster when it is younger and as it ages growth tends to slow down, resulting in an approximately negative exponential growth curve in raw ring width measurements (Dyer & Fritts, 1978). Before cross-dating, it is important to remove these growth trends to avoid autocorrelative effects on alignment. Samples were detrended by spline subtraction at a frequency response of 50%, at a wave length of $\frac{l}{2}$ years (Cook & Peters, 1981), to produce residual ring width measurements referred to as ring width “indices” (RWI).

For each sample, the Towai master sequence was rebuilt and detrended without the sequence containing the sample to be cross dated, to maintain independence between samples and master sequences and avoid influencing the likelihood of a match being found. Following Palmer et al. (2016), master sequences were detrended by log-transforming the Towai series then subtracting a 50-year spline with a frequency response of 20% from raw width measurements (Baillie & Pilcher, 1973; Bunn, 2010; Cook & Peters, 1981). Master sequences were built by taking the mean of spline detrended sequences.

For each combination of l , removed rings, and added rings (Table 4.1), sample generation was repeated 100 times, giving a total of 18,000 samples for cross-dating by Pearson correlation and DTW.

4.2.3 Alignment by Pearson Correlation and t -Test

The most common method of alignment in dendrochronology is by sliding-window Pearson correlation and Student’s t -Test statistics (Baillie & Pilcher, 1973). Fol-

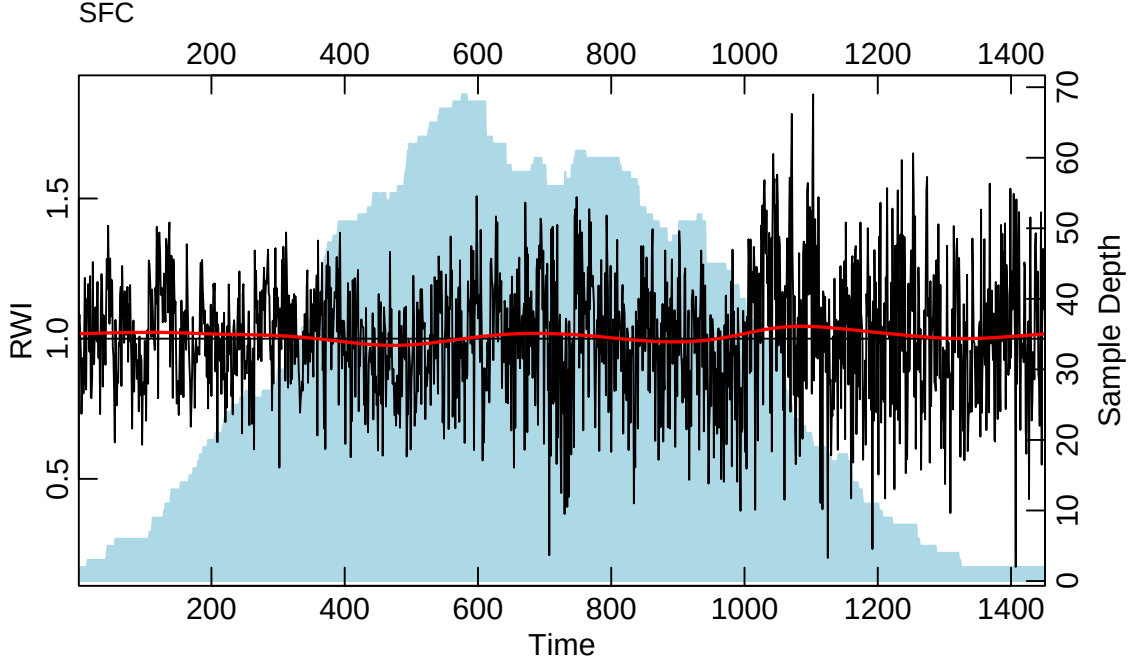


Figure 4.3: "Signal-free" detrended Towai master sequence and resulting ring width indices (RWI; black line; left y-axis), with 50-year spline overlay (frequency response = 20%; red line), and sample depth (light blue shading; right y-axis).

lowing [Hassan et al. \(2019\)](#), a brief overview of the method as implemented in this paper is given below.

Given a sample of ring width indices X of length N , and a master sequence of indices Y of length M (where $M \geq N$), cross-dating is performed by calculating the Pearson correlation coefficient (r) and Student's t -Test score (t) at all possible offsets of X to Y . The correlation coefficient of X and Y at a given offset (Δ) is given by

$$r(X, Y, \Delta) = \frac{\sum_{n=1}^N (X_n - \bar{X})(Y_{\Delta_n} - \bar{Y}_{\Delta})}{\sqrt{\sum_{n=1}^N (X_n - \bar{X})^2} \sqrt{\sum_{i=n}^N (Y_n - \bar{Y})^2}} \quad (4.1)$$

The Student's t -Test score is calculated using r and the length of the overlap (q) at the offset Δ :

$$t(X, Y, \Delta) = \frac{r\sqrt{q-2}}{\sqrt{1-r^2}} \quad (4.2)$$

As such, as r increases, so too does t . The t -Test allows meaningful comparisons of alignment across all possible offsets, as they may be of different lengths (q) due to

Table 4.2: Definitions of terms used in this study.

Term	Definition	Example
Series	A collection of multiple tree ring sequences, aligned to each other	Towai Farm dataset
Sequence	A timeseries of tree ring measurements, taken from a single tree	TOW002H
Subsequence	A subset of a sequence	TOW002H from index 100 to index 200
Sample	A timeseries of tree ring measurements for cross dating, with added or missing rings (in this study)	
Master sequence	A single timeseries of detrended and averaged tree ring measurements	Towai Farm master sequence

varying overlaps. The best alignment of X to Y is the offset that gives the highest t score, which in turn gives the most likely date for the first ring in the sequence.

The t -Test allows for meaningful comparisons between independent samples of different lengths when cross-dating in dendrochronology, giving a standardised measure of the quality of alignment. Typically, a confidence threshold of 3.5 is used as a minimum t -Test score in dendrochronology (Baillie & Pilcher, 1973). Alignments with a t -Test score less than 3.5 are typically considered inadequate for dating purposes.

4.2.4 Alignment by Dynamic Time Warping

In this section, an adjusted version of the DTW algorithm for dendrochronology is presented. A detailed description of the traditional DTW algorithm and its variations is provided in Chapter 3. For the sake of clarity, the core components of the DTW algorithm are reiterated here.

Simple symmetric DTW involves building a distance matrix (D) from the vectors X and Y , with each element given by the relationship (f) between X and Y at their varying indices:

$$D(n, m) = f(X_n, Y_m) \quad (4.3)$$

From D , a cumulative cost matrix (C) is defined:

$$C(n, m) = D(n, m) + \min \left\{ \begin{array}{l} C(n-1, m) \\ C(n, m-1) \\ C(n-1, m-1) \end{array} \right\} \quad (4.4)$$

4.2.4.1 Including Correlation in the DTW Distance Matrix (D)

In order to apply DTW to the problem of cross-dating in dendrochronology described above, several adaptations of the simple DTW algorithm were made. First, the function $f(X_n, Y_n)$ in Equation (4.3) was adjusted to include a measure of correlation. This requires incorporating multiple (v) values of X and Y , at neighbouring indices to n and m . Here, v was set to 5, meaning 4 additional values of X and Y are included to build D , at indices greater than n and m :

$$D(n, m) = f(a, b) \quad (4.5)$$

where

$$a = \begin{bmatrix} X_n \\ \vdots \\ X_{n+v-1} \end{bmatrix} \text{ and } b = \begin{bmatrix} Y_m \\ \vdots \\ Y_{m+v-1} \end{bmatrix} \quad (4.6)$$

Values of D where $n \geq N - v$ and $m \geq M - v$ were set to 0. The Pearson's correlation coefficient (r) was used to redefine the function f , such that:

$$D(n, m) = \frac{\sum_{w=1}^v (a_w - \bar{a})(b_w - \bar{b})}{\sqrt{\sum_{w=1}^v (a_w - \bar{a})^2} \sqrt{\sum_{w=1}^v (b_w - \bar{b})^2}} \quad (4.7)$$

Given that higher values for r indicate closer similarity, the minimising function in Equation (4.4) was adjusted to a maximising one (Equation (4.8)).

4.2.4.2 Constraining Alignments to Linear Solutions

In dendrochronology, it is reasonable to assume that solutions to the alignment problem will be linear, except for where missing or added rings are detected. In

other words, each year (ring) in a sample can be expected to represent one year in the master sequence, and only where missing or added rings are very likely should the solution deviate from a linear alignment between the subsequence and master sequence. For this reason, a non-linear alignment penalty z was added to Equation (4.4), similar to those commonly applied in DTW as step patterns (see Chapter 3, Giorgino, 2009). This parameter, here set to 0.7, helps to constrain alignment to linear solutions and avoid complex warping paths across the matrix D :

$$C(n, m) = D(n, m) + \max \left\{ \begin{array}{l} C(n-1, m) \cdot z \\ C(n, m-1) \cdot z \\ C(n-1, m-1) \end{array} \right\} \quad (4.8)$$

4.2.4.3 DTW and Adjusted t -Test Statistic

Unlike traditional DTW, the final DTW distance of the comparison between X and Y in the adjusted DTW algorithm described is given by $C(M-v, N-v)$, given the constraints of Equations (4.5), (4.6), and (4.7). This distance, as defined by Equation (4.8), is a *maximised* score, meaning that, counterintuitively, two signals with a higher DTW distance are *more* similar than those with a lower score.

In order to make DTW distances comparable across varying lengths of X and Y , the DTW score is divided by $N(1+k)-v$ to produce a normalised DTW distance, α :

$$\alpha = \frac{C(N-v, M-v)}{N(1+k)-v} \quad (4.9)$$

where k is a “window size” parameter, described in Section 4.2.4.4.

This method is similar in approach to path normalisation by standard DTW methods (Giorgino, 2009). Conveniently, the DTW score at $C(N-v, M-v)$ is unlikely to exceed N , given that each value of M cannot be greater than 1, and that z constrains alignments to linear or near-linear warping solutions. As such, dividing by $N(1+k)-v$ approximately scales DTW scores between 0 and 1, in turn allowing for an adjusted t -Test score (t_{DTW}) comparable to Equation (4.2):

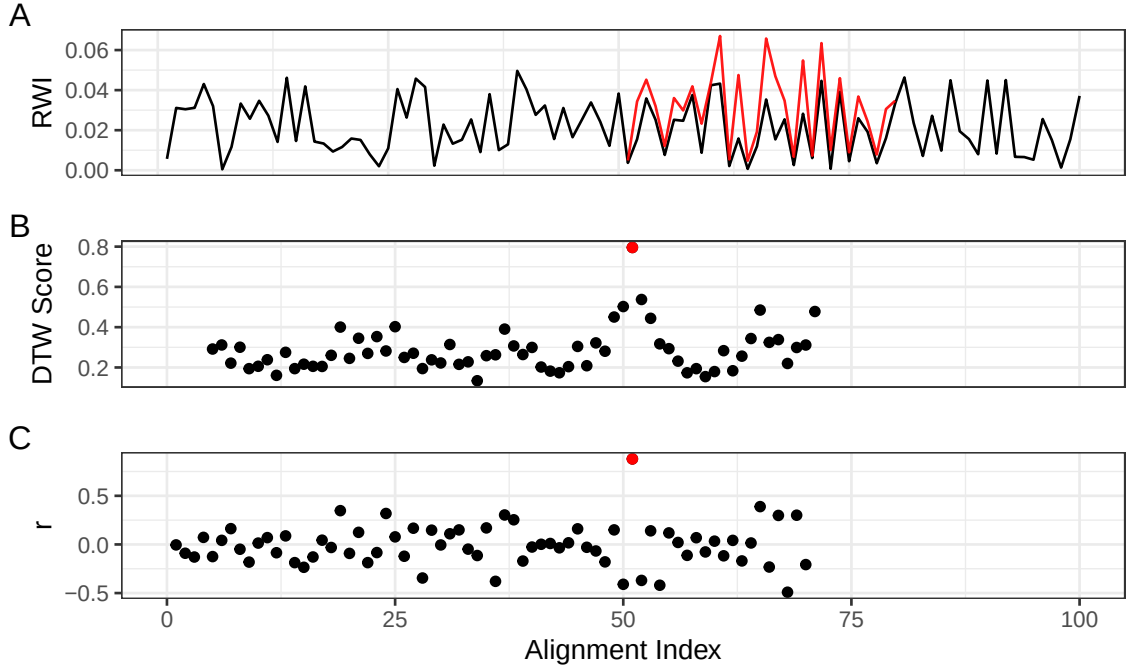


Figure 4.4: An example alignment problem and its solution by DTW and Pearson correlation. In Panel A, the reference vector (black line) and query (red line) are shown, with the query positioned at the correct index (i.e., the solution). Panel B shows the scores of sliding window alignment by DTW, the maximum of which gives the solution. Panel C is as for Panel B, but for Pearson correlation.

$$t_{DTW}(X, Y, \Delta) = \frac{\alpha \sqrt{N-2}}{\sqrt{1-\alpha^2}} \quad (4.10)$$

4.2.4.4 DTW Distance and Sliding Window Approach

A sliding window approach was used to find the highest values of α across all possible alignments (Figure 4.4). As described above in Section 4.2.3, in cross-dating by Pearson correlation an alignment match is searched for across the master sequence, by calculating r at all possible offsets of X to Y . A similar method is applied here using the DTW algorithm: α is calculated at all possible offsets (Δ).

However, in order for the DTW to be flexible enough for dendrochronology, and particularly to accommodate added rings, it is important that the length of each subset of Y ($|Y_\Delta|$) is *longer* than the length of X . This is not the case for Pearson correlation, where vectors X and Y_Δ must be of equal length (Equation (4.1)).

As such, for each Δ , a subsequence of Y of length $N(1 + k)$ is selected and considered for alignment, where N is equal to the length of X , and k is a “window size” parameter that extends each subsequence of Y , specifically to allow cases of added rings. In this study, k was fixed at 0.1, meaning that each subsequence of Y compared by sliding window DTW is 10% longer than X . For example, for a subsequence X with length 20, the length of each subsequence of Y considered for alignment is 22.

The index of the best alignment of X to Y by DTW, similar to alignment by Pearson Correlation, is given by the offset (Δ) with the highest α score (Equation (4.9)).

4.2.5 Assessing Accuracy

DTW and Pearson correlation methods were compared by their proportions of correct alignments (i.e., those where the correct location of the sample within the master sequence was identified). Given that rings were added to and removed from samples in the experimental design after selection from sequences, it is possible that “correct” alignments by DTW or by correlation were inaccurate in terms of their alignment to master sequence indices by up to 5 years. Therefore, a “correct” match was defined as one within 5 years of the original date.

As described in Section 4.2.3, a t -Test score of 3.5 is typically used as a threshold for an “acceptable” date (Baillie & Pilcher, 1973). The threshold of 3.5 is somewhat arbitrary, but has proven to be effective for most cases. However, as outlined by Wigley et al. (1987), there can be cases where an alignment may be correct, but the t -Test score falls below this threshold, or where an incorrect alignment’s t -Test score is above the threshold. Rather than consider alignments with t -Test scores less than 3.5 incorrect matches, we assess their accuracy as described above, and consider their t -Test scores independently.

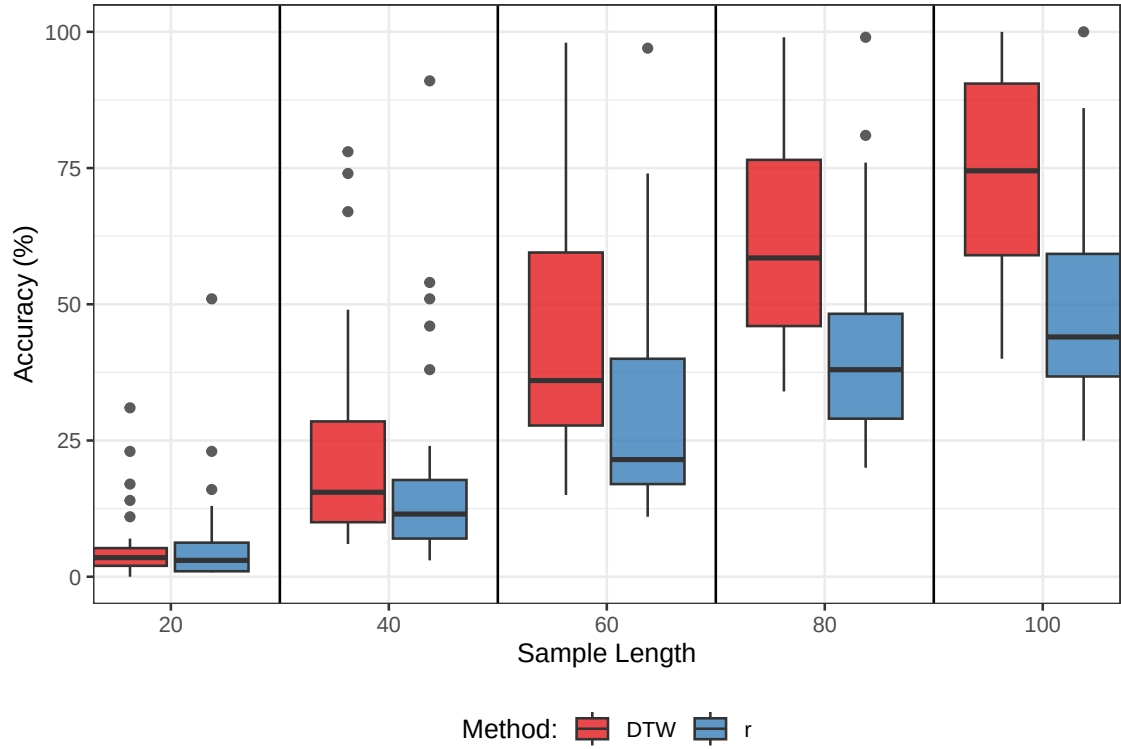


Figure 4.5: Boxplots summarising the accuracy of alignments by DTW (red) and Pearson correlation (blue) at varying sequence lengths between 20 and 100 (x-axis), for all numbers of rings added or removed.

4.3 Results

The effectiveness of Pearson correlation and alignment by DTW for varying sequence lengths is summarised in boxplots in Figure 4.5, showing the accuracy (% correct matches) of both methods for all numbers of rings added and removed. Generally, DTW outperformed Pearson correlation, particularly for higher sample lengths, but was less effective for shorter samples. The average accuracy over all sequence lengths and combinations of added and removed rings was higher for DTW ($M = 43\%$, $SD = 31\%$) than for correlation ($M = 29\%$, $SD = 23\%$).

Figure 4.6B shows the distribution of samples produced by the two-step random sampling strategy described in Section 4.2.2. Although sampling from the Towai series was randomised, the distribution of samples across indices is not even, as the number of sequences available for sampling (i.e., series depth) is lower at the beginning and end of the series. Figure 4.6B shows that variability in series depth

closely follows the sampling distribution over the Towai series' indices, suggesting that there were no biases in the sampling strategy.

The distribution of correct alignments for DTW and correlation methods over the Towai indices are presented in Figure 4.6C, with both largely following the sampling distribution and Towai series' depth. However, some deviation is apparent for Pearson correlation: between indices ~ 600 and 650 , a disproportionate reduction in correct alignments is shown, corresponding to the beginning of GS-1 and the period before it. Figure 4.6C shows the densities of incorrect alignments over their incorrect indices (i.e., the location that either method incorrectly cross dated the sample to). Incorrect bias to the later parts of the series are apparent for alignment by Pearson correlation. The distribution of incorrect matches by DTW is more even over the series, with no clear biases in alignment to any particular part of the sequence.

Figure 4.7 shows the % accuracy of alignment by DTW (top row) and Pearson correlation (bottom row), for varying sequence lengths (columns), and for varying numbers of added and removed rings (x- and y-axes, respectively). For both DTW and correlation methods, accuracy generally decreased when rings were added or removed, and overall accuracy improved as sequence length was increased. For shorter sequence lengths ($l \leq 40$), accuracy by Pearson correlation was higher than DTW when no rings were added or removed, but was comparable or lower than DTW when rings were added or removed. For longer sequence lengths ($l \geq 60$), DTW consistently outperformed correlation, and was equivalent to Pearson correlation when no rings were added or removed.

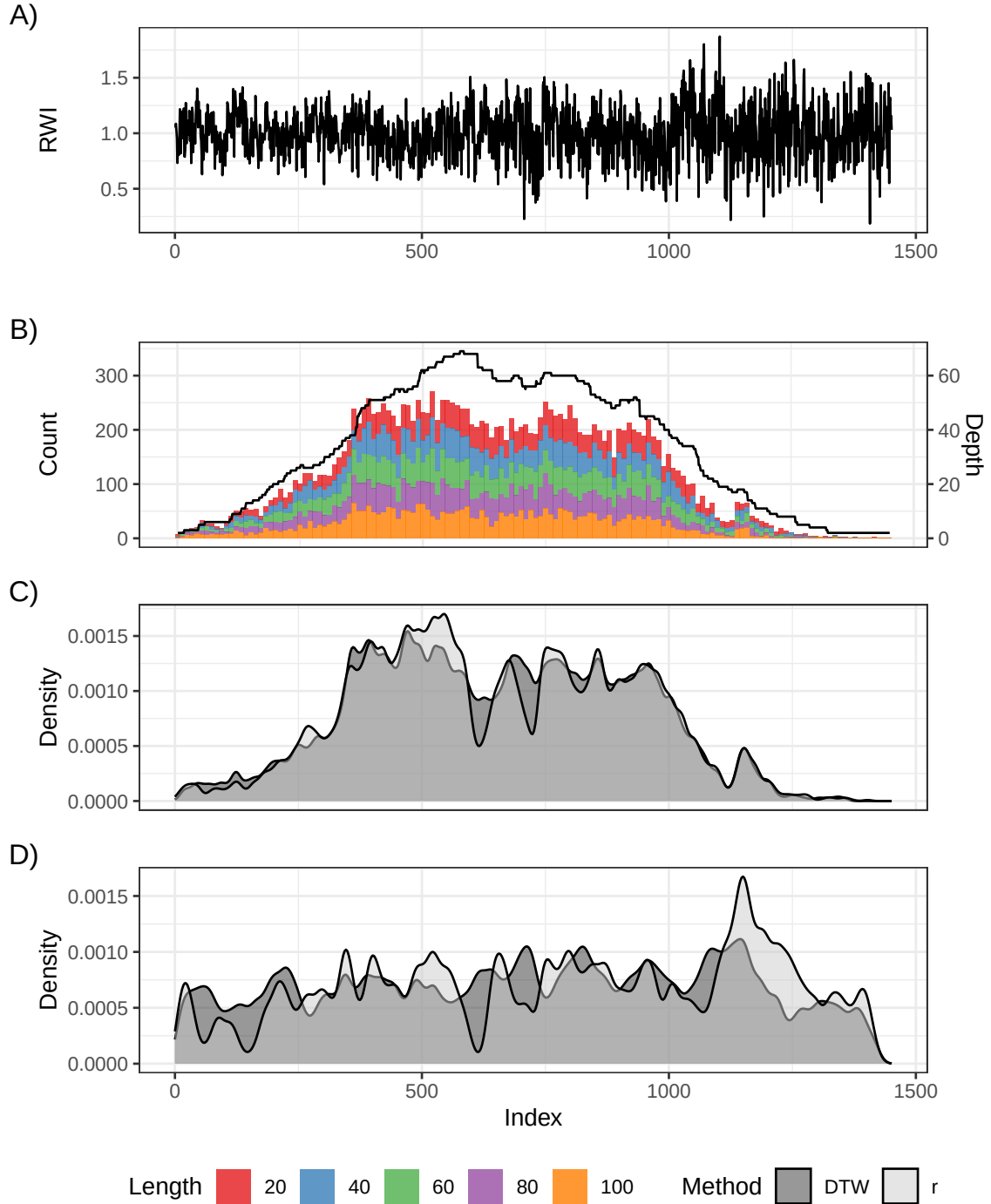


Figure 4.6: Sampling distribution and frequencies of correct and incorrect matches over the Towai series. Panel A shows the Towai master sequence after spline detrending, using all sequences. Panel B shows the Towai series depth (black line) and sampling distribution over the series' indices, coloured by sequence length (bin width = 10). Panel C shows the distribution of correct sequence matches for DTW (dark grey) and Pearson correlation (light grey) alignment methods. Panel D shows the distribution of incorrectly matched indices for DTW and Pearson correlation, showing an alignment bias for Pearson correlation around index 1100.

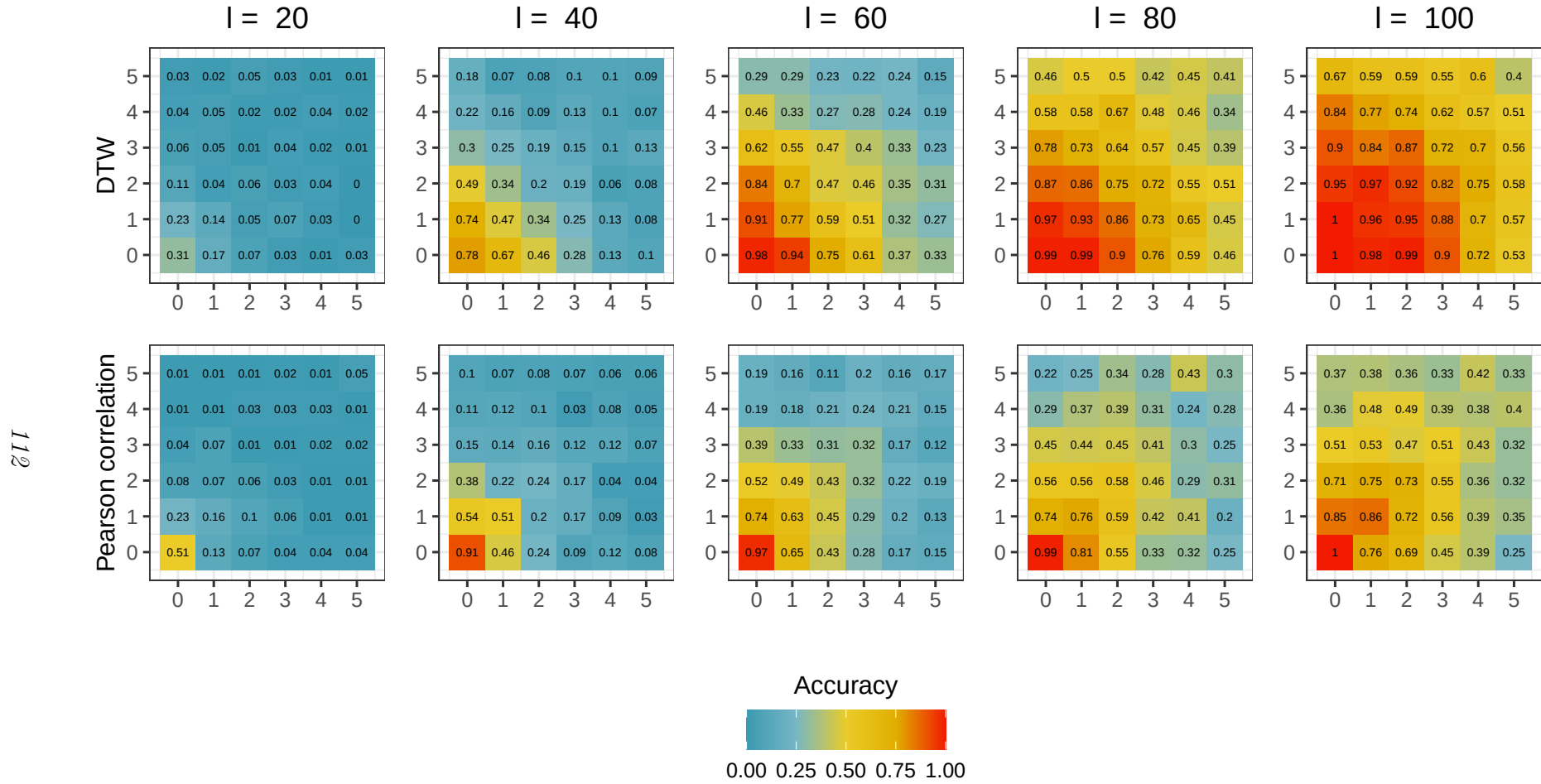


Figure 4.7: Accuracy of alignments for varying sequence lengths by DTW (top row) and by Pearson correlation (bottom row). Values in cells show the proportion of correctly matched samples for rings added and rings removed. A correct alignment is defined as being within 5 indices of the original sample location. Rings added are shown on horizontal axes (from 0 to 5), rings removed are shown on vertical axes.

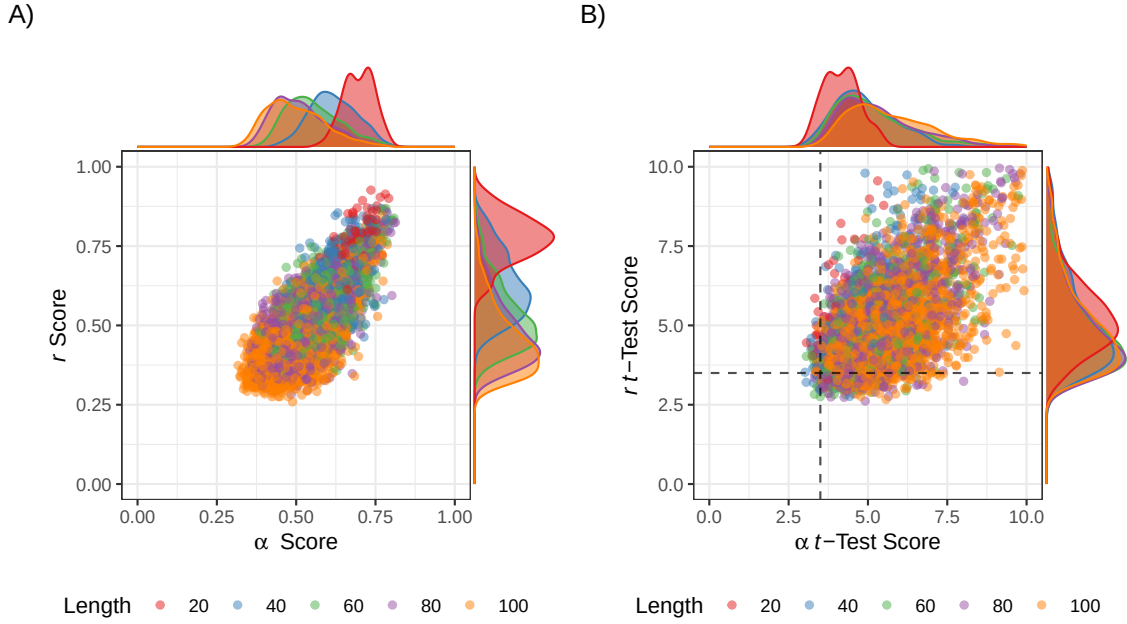


Figure 4.8: Paired DTW (α) and Pearson correlation (r) scores for correct alignments. Panel A shows the relationship between α and r for cases where both methods found the correct solution, coloured by sequence length. A positive relationship between α and r suggests that for cases where Pearson correlation is high, DTW scores are also typically high. Panel B shows the same paired scores after adjusting for sample length by t -Test. Densities of α , r , and t -Test scores are shown along the upper and right axes of both panels, showing less clear separation after t -Test adjustment.

The relationship between correct DTW alignment scores (α) and correct Pearson correlation alignment scores, along with their densities, are shown in Figures 4.8 and 4.9. Figure 4.8A shows the relationship between scores for correct matches by both methods, and Figure 4.8B shows the scores adjusted by t -Test. A positive correlation is apparent between both methods ($r = 0.84$, $p < 0.005$), meaning that correct samples with a higher DTW score also had higher Pearson correlation coefficients. Scores for correct alignment by DTW and correlation were generally higher for shorter sample lengths. When accounting for sequence length by t -Test score, the distinction between varying sequence lengths is less clear, which is to be expected. For correlation, 37.4% of samples produced t -Test values less than the threshold of 3.5, compared to 19.7% for cross-dating by DTW (dashed lines in Figure 4.8B).

Figure 4.9 shows the relationship between incorrect alignments by both methods. Again, a positive correlation is apparent, with much more separation between den-

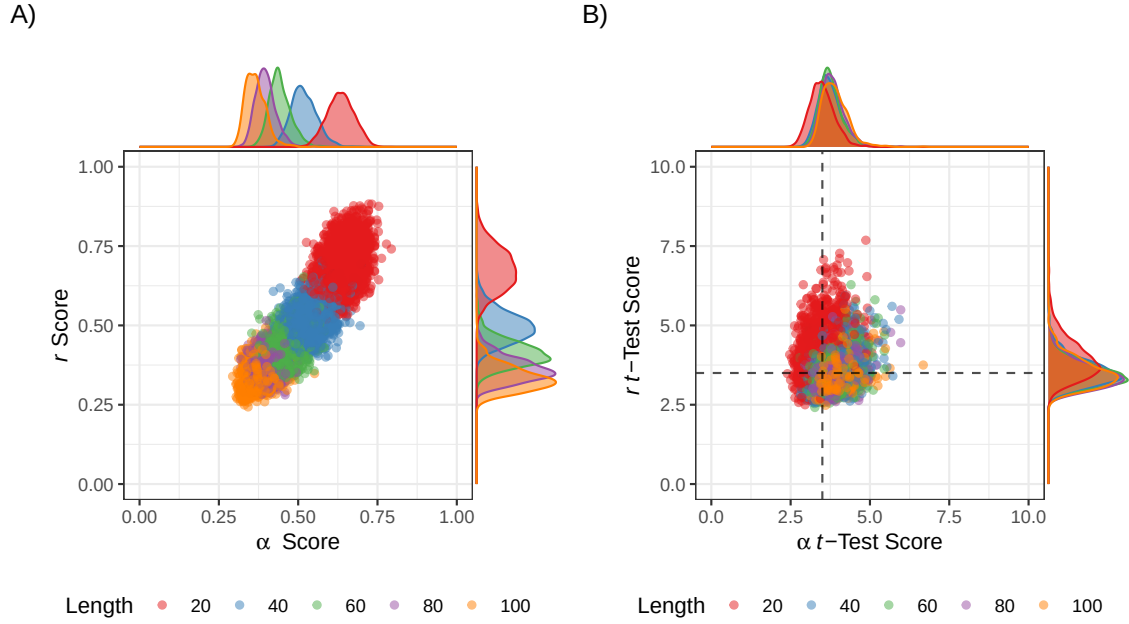


Figure 4.9: Paired DTW (α) and Pearson correlation (r) scores for incorrect alignments. Panel A shows the relationship between α and r for cases where both methods did not find the correct solution, coloured by sequence length, as for Figure 3.8. Panel B shows the same paired scores after adjusting for sample length by t -Test. Densities of α , r , and t -Test scores are shown along the upper and right axes of both panels, showing less clear separation after t -Test adjustment.

sities (Figure 4.9A margins). Accounting for sample length by t -Test (Figure 4.9B) again shows less separation. Figure 4.9B shows that many incorrect alignments by both methods passed the t -Test threshold of 3.5.

4.4 Discussion

The experiment presented in this chapter explores an alternative method for cross-dating in dendrochronology, using an adjusted DTW algorithm as a distance measure between samples and a master sequence. The experiment design assessed the effectiveness of DTW for cross-dating and alignment, specifically for cases where missing or rings are present in samples. A dataset of tree ring sequences was repeatedly sampled to produce samples for cross-dating containing varying numbers of added and missing rings (Palmer et al., 2016). An adapted version of the DTW algorithm that uses a correlation-based similarity metric to build cost matrices was then applied to samples from the series for cross-dating. The accuracy of cross-dating by DTW was assessed and compared to Pearson correlation, the latter being the dominant method in dendrochronology.

For cases where sample lengths were greater than or equal to 60, and no rings were added to or removed from samples (a typical cross-dating scenario), the accuracy of alignment by DTW was equivalent to Pearson correlation (Figure 4.7). For shorter samples with no added or missing rings, Pearson correlation was more accurate than DTW, but only 51% of samples were correctly matched by Pearson correlation for samples of length 20. Where rings were added or removed from samples, DTW was consistently more accurate than Pearson correlation for lengths over or equal to 60, but both methods were ineffective for shorter lengths with added or missing rings. These results suggest that:

1. Pearson correlation can be problematic for cases where added or missing rings are present (Figure 4.7).
2. For sample lengths above ~ 60 rings, DTW has valuable potential in cross-dating in dendrochronology, particularly for cases where no strong match is found between a master sequence and sample by Pearson correlation (Figure 4.5).

3. A t -Test threshold of 3.5 may be an ineffective measure for the “quality” of alignment by Pearson correlation, as many correct matches by Pearson correlation had t values less than 3.5. This is less of an issue for DTW (Figure 4.8).

The implications of the sampling strategy, parameters, and accuracy metrics designed in the experiment, as well as the broader results are discussed in sections below. Finally, the shortcomings of both methods and recommendations for future research in this field are outlined.

4.4.1 Distribution of Samples and Alignments

The distribution of samples as built from the [Palmer et al. \(2016\)](#) dataset closely follows the sample depth of the Towai series, as shown in Figure 4.6B. In practice, it is unlikely that the beginning or end of the Towai series would be considered for cross-dating in dendrochronology; [Palmer et al. \(2016\)](#) suggest that indices before ~ 150 and after ~ 1250 in the Towai series could be unreliable given their low sample depth and corresponding EPS values (Figure 4.10). A revised experimental design accounting for uncertainty in these areas, by removing the possibility of sampling from and alignment to regions with an EPS below 0.85, for example, would likely affect results and improve the accuracy of both methods. Incorrect alignments were made by both Pearson correlation and DTW to the beginning and end of the Towai series (Figure 4.6 D), and it is possible that these alignments would have been correct with a revised master sequence scheme using an EPS threshold. However, [Buras \(2017\)](#) argues against the use of EPS in altogether, suggesting its application and adoption in dendrochronology is erroneous. Exploring other metrics alongside such an experiment, such as subsample signal strength ([Buras, 2017](#)), might therefore be worthwhile.

Comparing densities of correct matches between Pearson correlation and DTW, shown in Figures 4.6C and 4.6D, reveals significant differences their in distributions, particularly between indexes ~ 600 and 700 . This period in the Towai dataset

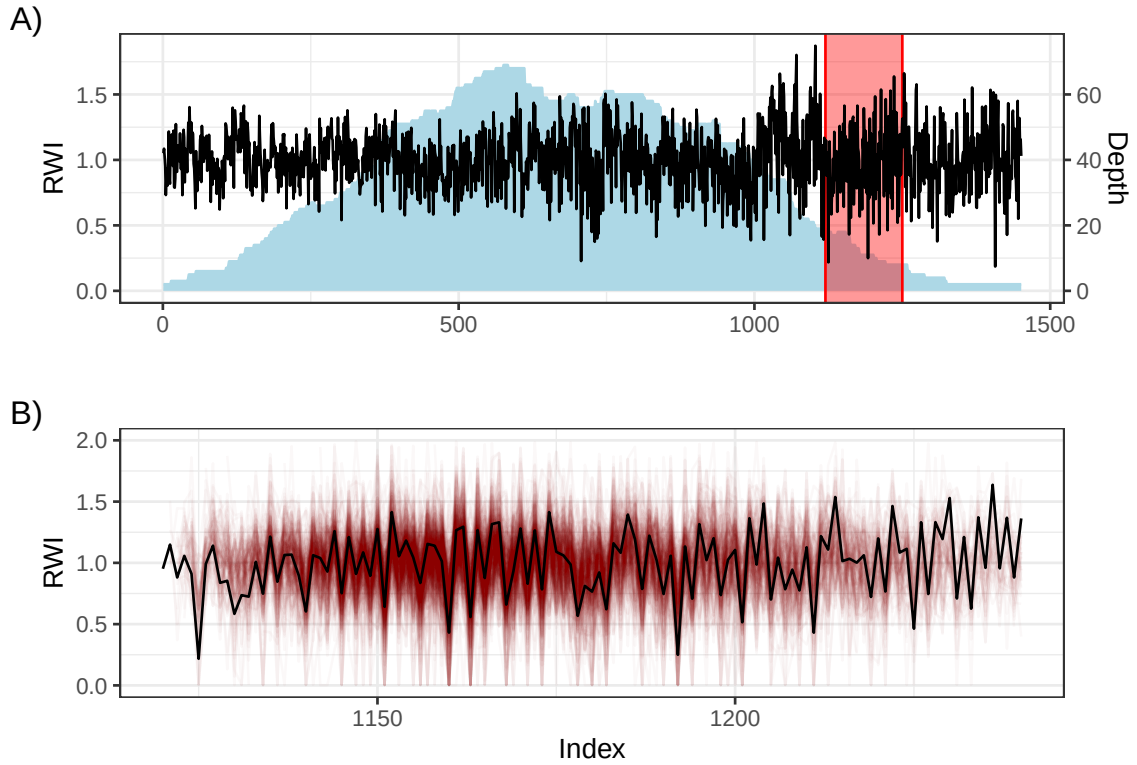


Figure 4.10: The Towai master sequence from indices 1120 to 1240, where some bias is apparent for alignment by Pearson correlation. Panel A shows the Towai master sequence (black line, left y-axis), series depth (blue shading, right y-axis), and highlights the region between indices 1120 and 1240 (red vertical lines and shaded region). Panel B shows indices 1120 to 1240 of the Towai master sequence. Incorrect alignments by Pearson correlation are shown in red ($N = 761$), with the master sequence overlain in black.

approximately spans the ring width downturn corresponding to GS-1 in Greenland ice core records, and the period of climate “flickering” immediately preceding it (Palmer et al., 2016; Rasmussen et al., 2014).

Alignment by Pearson correlation shows a reduction in both correct and incorrect matches for GS-1 and the surrounding period, suggesting that the method holds some inherent bias against alignment to the flickering and downturn. The distribution of incorrect alignments for Pearson correlation (light grey in Figure 4.6D) also shows that a disproportionate number of samples were incorrectly aligned to later indexes in the master sequence (approximately between indexes 1120 and 1240). Examining the two-dimensional densities of incorrect alignments by Pearson correlation, over original and misaligned indices (Figure 4.11), confirms that the

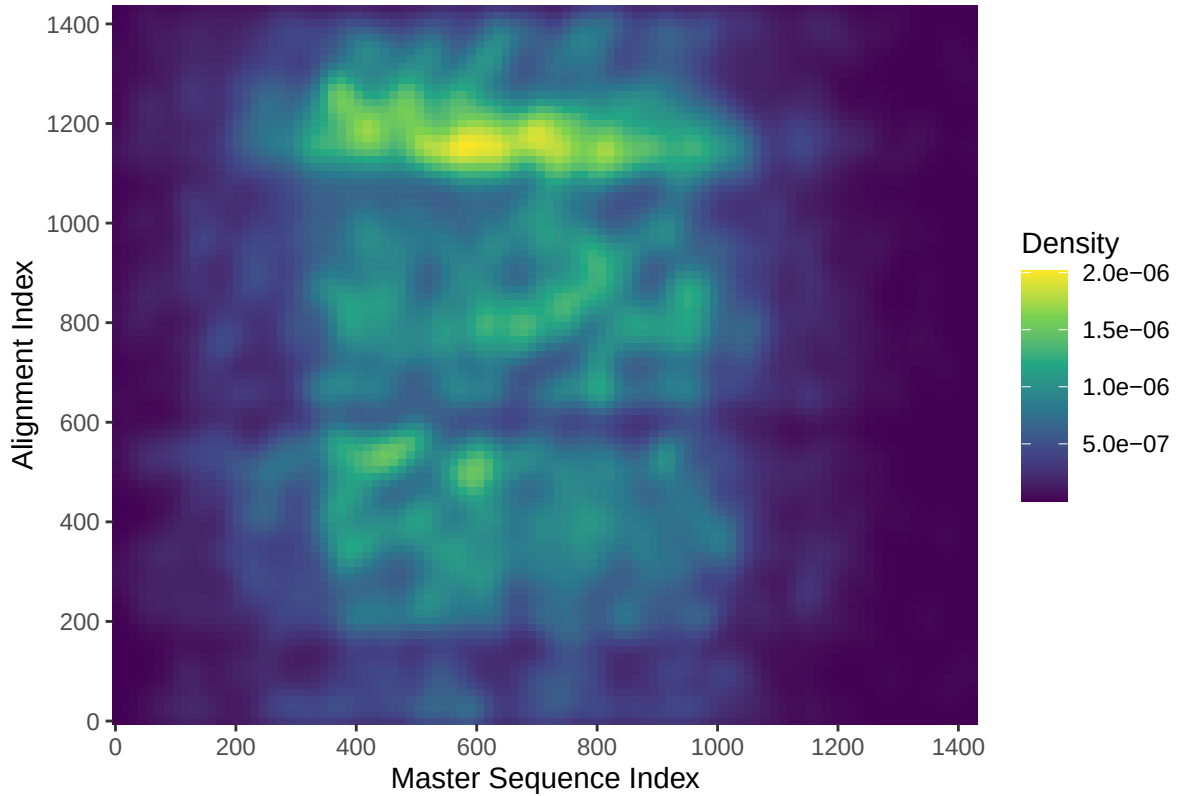


Figure 4.11: Two-dimensional density plot of incorrect alignments by Pearson correlation, showing the original Towai indices (x-axis) and the alignment indices (i.e., where samples were incorrectly aligned; y-axis). Some bias to alignment is apparent around index 1200, with higher density in this region (y-axis).

incorrect matches from GS-1 (approximate master sequence index 600, x-axis) have some bias to this region of the Towai series (approximate alignment index 1150, y-axis). However, a broader bias to index ~ 1150 is apparent for many other regions of incorrect alignments by Pearson correlation shown in Figure 4.11.

Inspecting the section of the master sequence around index 1150 in more detail (Figure 4.10) reveals no obvious reason for such a bias, such as increasing or decreasing trends not removed by spline detrending that might “attract” autocorrelative alignment. The magnitude of fluctuations in RWI is somewhat greater than other periods, but is not excessive. Some distinct and dramatic fluctuations appear to be consistently aligned to by Pearson correlation (e.g., $\sim$ index 1160, Figure 4.10B). However, these factors alone do not explain such a bias, as they are not particularly unusual within the whole Towai series.

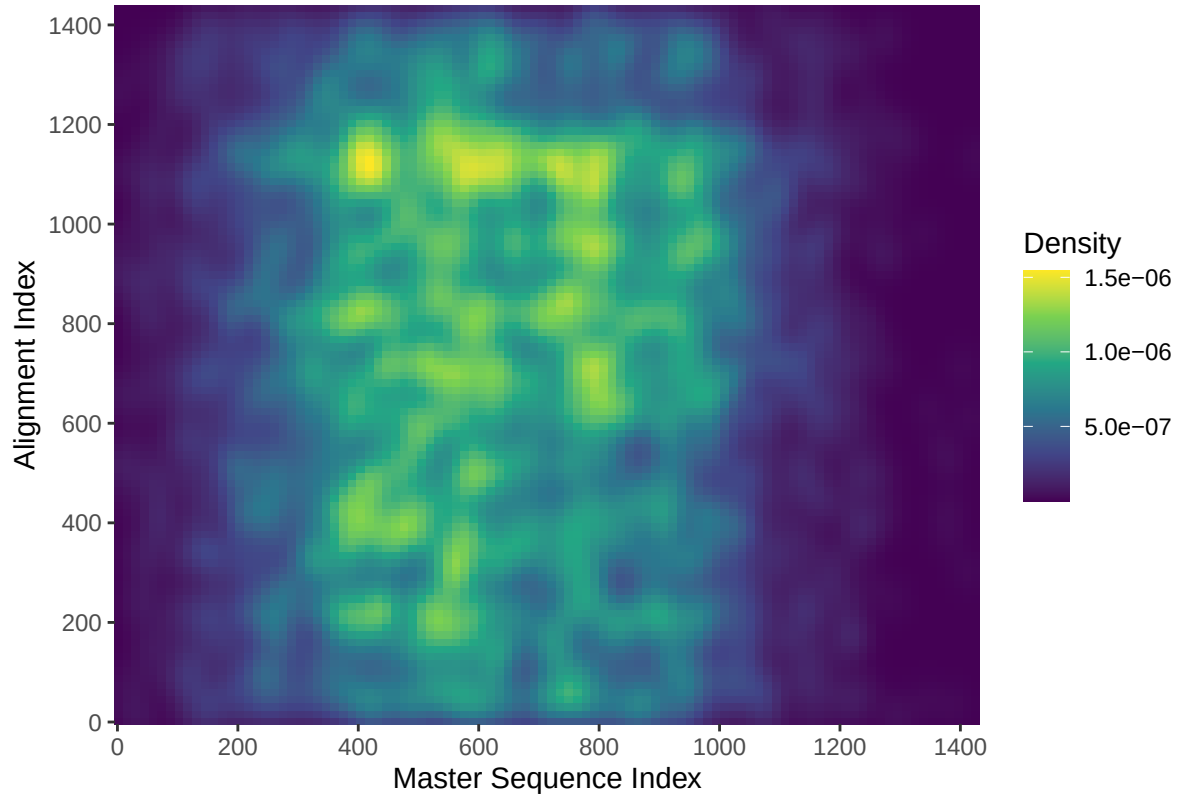


Figure 4.12: Two-dimensional density plot for incorrect alignments by DTW, showing original Towai indices (x-axis) and the alignment index (i.e., where samples were aligned; y-axis).

DTW, by contrast, does not seem to suffer as dramatically from this unexplained bias. The two-dimensional densities of incorrect alignment by DTW show a more even distribution over indices, although some concentration around the 1150 indices is also apparent (Figure 4.12). Repeating the experiment presented here using different and multiple series might reveal whether this phenomenon is distinct to the Towai dataset, or whether alignment by Pearson correlation suffers similar biases when applied in other contexts. If, for example, high magnitude variability or distinct patterns in RWI values consistently attract misalignment by Pearson correlation for other datasets, while DTW is consistently more evenly distributed in its incorrect matches, then there might be further justification for applying DTW in dendrochronology.

Some improvements might be made to the sampling strategy employed in this study. In the Towai series, many sequences are measured from separate cores taken

from the same tree (Figure 4.2). Although individual sequences (e.g., TOW101A) were removed from the series before master sequence building, sequences from the same tree were not (e.g., TOW101B, TOW101C, etc.). This may have had an effect on the likelihood of correctly aligning the sample; in effect, the samples and master sequences in each iteration of the experiment were not independent. Correcting for this by also removing sequences would be relatively straightforward, but this would dramatically affect series depth and EPS values for the Towai series. This adjustment in combination with an EPS threshold or similar, as discussed above, might alleviate this.

Interseries correlation (correlation between each sequence and the master sequence) is relatively high for the Towai series ($M = 0.68$, $SD = 0.07$). Figure 4.13 shows clear, linear, and positive relationships between interseries correlation and the proportion of correct matches (i.e., % accuracy) for each sequence and method. This suggests that, rather unsurprisingly, both methods are more accurate when samples originated from sequences that agreed well with the master sequence. For DTW, this relationship has less variance, shown by a higher R^2 value (DTW = 0.7, Pearson correlation = 0.54). Again, applying a similarly structured experiment to other tree ring datasets, perhaps that are not so exceptionally well-correlated and that contain more independent measurements (i.e., less cores from the same trees), might further illuminate the potential for applying DTW to cross-dating in dendrochronology.

4.4.2 Parameter Sensitivity and Variable Sample Lengths

For sample lengths greater than or equal to 60, DTW clearly outperforms Pearson correlation in cases where rings are added or removed from samples (Figure 4.7), and is comparable to Pearson correlation where no rings are added or removed. The experiment design and its results show that DTW is very resilient to the problem of added or missing rings in dendrochronology. If long sequences can be obtained for cross-dating, DTW performs as well as Pearson correlation in aligning samples to a master sequence.

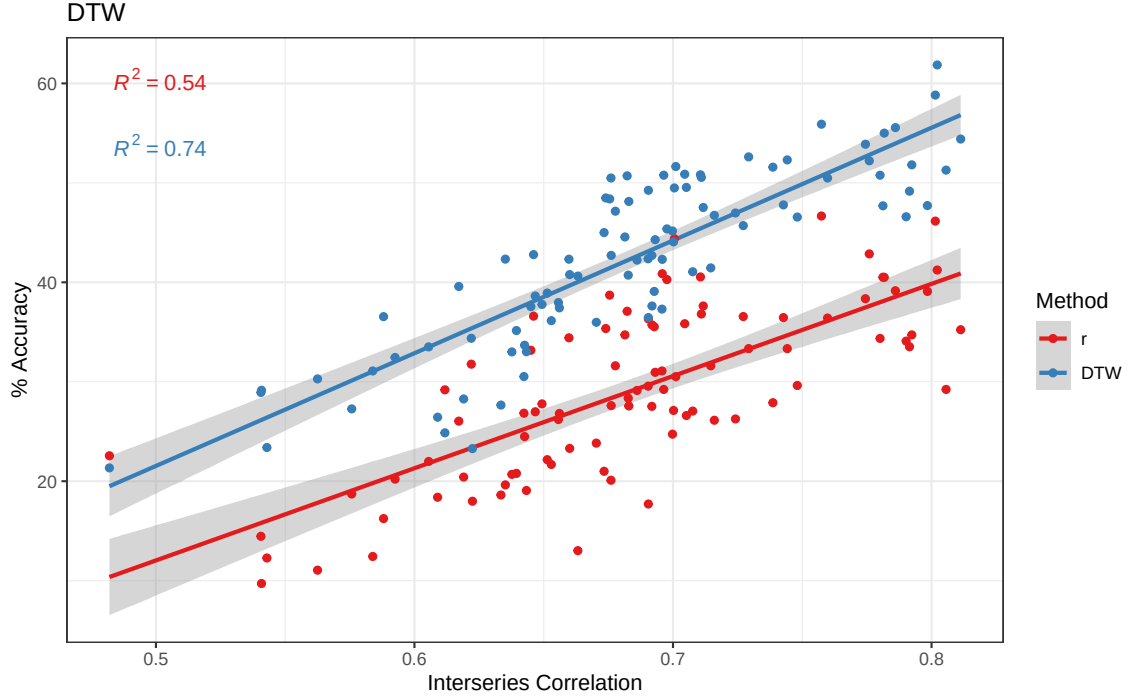


Figure 4.13: The relationship between interseries correlation (x-axis) and % accuracy (y-axis) for each sequence in the Towaii series (e.g., TOW101A), for DTW (blue) and Pearson correlation (red), with linear regression and R^2 values. Generally, both alignment methods were more accurate where interseries correlation was higher.

Poorer performance by DTW for lower sequence lengths ($l = 40$ and $l = 20$) might be attributed to choices for a number of parameters set in Table 4.1. First, the DTW window size (k) was set to 0.1, meaning that for shorter sequence lengths, alignment by DTW is unlikely to find a correct alignment for cases where the number of rings added is greater than $|Y_\Delta|(1 + k)$. For example, where $k = 0.1$ and $n = 20$, DTW is unlikely to find a correct match where the number of rings added is greater than 2. Because in this example $|Y_\Delta| = 22$, the number of likely “insertions” found by DTW is constrained to 2, unless very high correlations are found for insertions while building the matrix D (Equation (4.8)). Related to this, DTW appears to perform better for missing rings than for added rings. This might also be attributed to too low a k value, given that deletions (missing rings) are not constrained by the length of Y_Δ .

Second, setting the parameter z to a lower value might allow a greater degree of non-linearity in warping paths, which would in turn allow for greater numbers

of insertions, alleviating the affect of a low k value. However, in experimental testing, lowering z frequently resulted in excessively complex warping paths across the matrix C , with an overall lowering of accuracy.

Third, the parameter v also has an effect on alignment by DTW, and this effect becomes greater as sample length decreases. If v is constant across all sample lengths (in this study set to 5), a greater proportion of the matrix D is set to zero for shorter sequence lengths (Equation (4.5)), and a greater proportion of the information in the query is lost where sequence lengths are lower. For a sample of length 20, around a 25% of the matrix D is set to zero, whereas for a sample of length 100, only 5% of D is set to 0. This is the most likely reason for poor performance by DTW for $l = 20$, even for cases where no rings are added or removed (Figure 4.7).

The sample index of removed or added rings also has an effect on cross-dating outcomes by Pearson correlation. Depending on the relative index of missing or added rings in X , alignment by Pearson correlation will be more or less susceptible to incorrect alignments in Y . If removed or added rings (or both) are located at the beginning or end of X , correlation is more likely to find a correct match, because a greater *continuous* portion of the sample is preserved for cross-dating. If a ring is added or removed at the centre of X , only half of the sample is useful in finding a correct alignment by Pearson correlation.

This effect is summarised in Table 4.3. Using the sample and master sequence generated for the example presented in Figure 4.4 ($|X| = 30$), Pearson correlation and DTW scores are shown in Table 4.3, for a single removed ring at the 5th (near-beginning), 15th (middle), and 25th (near-end) indices. Pearson correlation is unable to find the correct alignment (index 51) for the case where the middle (15th) ring is removed from the sample. Where sample lengths are shorter, this effect will be more pronounced.

In practice, sample lengths of less than 50 are unlikely to be considered for cross-dating (Wigley et al., 1987), and results here confirm that a minimum sample length of 50 is an appropriate threshold. It is also very unlikely that any sample

Table 4.3: Example alignment problem for Pearson correlation and DTW, as shown in Figure 4.4, with one ring removed at the 5th (near-beginning), 15th (middle), and 25th (near-end) indices. Pearson correlation (r) and associated statistics (p , t -Test) are shown, along with DTW distance (C), normalised DTW distance (α), and the DTW adjusted t -Test score. Alignment indices (i.e., solutions) for both methods are also shown. The correct alignment is to index 51. Where the middle (15th) index is removed, Pearson correlation fails to find the correct solution, with a low r value and failed t -Test. DTW finds the correct solution (within one index) for all removed rings, passing the t -Test threshold.

Index	Pearson correlation				DTW			
	r	p	t -Test	Alignment	C	α	t -Test	Alignment
	0.88	< 0.005	9.77	51	21.79	0.78	6.56	52
5	0.83	< 0.005	7.77	52	19.80	0.74	5.65	53
15	0.48	0.008	2.85	38	16.69	0.62	4.11	52
25	0.79	< 0.005	6.79	51	20.01	0.74	5.78	52

considered for cross-dating will contain five (or even two) missing and added rings. In this sense, the experimental design of this chapter demonstrates the theoretical, rather than practical, applications of DTW in dendrochronology, and its advantages over Pearson correlation.

4.4.3 Accuracy Metrics

One of the key advantages of dendrochronology over other dating methods is its precision, with very little or no error attached to dates where the master sequence chronology is well constrained. Due to the accuracy threshold of 10 years applied in this experiment, the DTW algorithm presented carries an effective uncertainty of ± 5 years. The only measure of confidence in alignments is the t -Test score, and α value used to produce it. Further development of the DTW algorithm for “real world” application would require resolving a clear one-year bias in its alignments (Figure 4.14), as well as using a more stringent accuracy metric.

If further developing the algorithm for regular deployment to dendrochronological problems, it would be fairly straightforward to use the warping path across the matrix defined in Equation (4.8) to detect and highlight missing or added

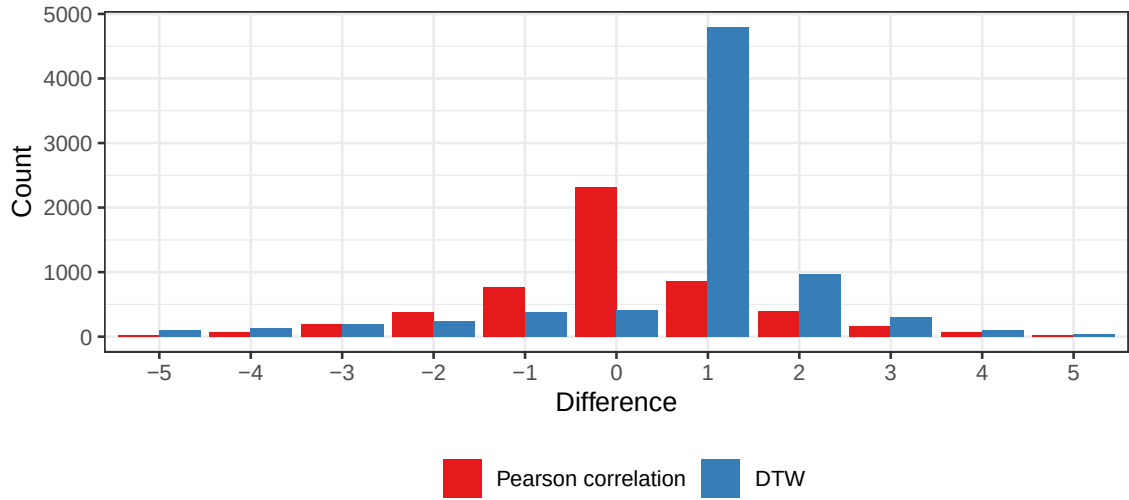


Figure 4.14: Distribution of differences between original index (i.e., the solution) and alignments by DTW and Pearson correlation, for correct matches (alignment falling within the accuracy threshold of 10 years).

rings. Non-linearities in the path (insertions or deletions) can easily be used to infer missing or added rings, with their indices used to adjust the final alignment. Automated detection of and adjustment for missing or added rings could easily be integrated into the DTW algorithm, as well as existing dendrochronology programs. This would avoid the difficulty of having to repeatedly run analyses at shorter segment lengths to detect missing or added ring indices by visual inspection, such as in COFECHA or dplR (Bunn, 2008; Grissino-Mayer, 2001).

An adjusted t -Test statistic was calculated and used to compare α scores to t -Tests for Pearson correlation. Figure 4.15 shows that the t -Test fails to separate correct and incorrect matches adequately for either method, as highlighted previously (Wigley et al., 1987). The t -Test threshold of 3.5, by extension, is confirmed as arbitrary and inadequate: for both alignment methods, high proportions of incorrect matches pass the t -Test threshold (68% of incorrect matches for DTW; 52% of incorrect matches for Pearson correlation); for Pearson correlation a considerable number of correct alignments' t -Test values fall below the threshold (13%); while 97% of correct alignments by DTW pass the t -Test threshold. This confirms that alternative methods should be considered for quantifying the confidence in alignments, such as probabilistic methods (Hassan et al., 2019).

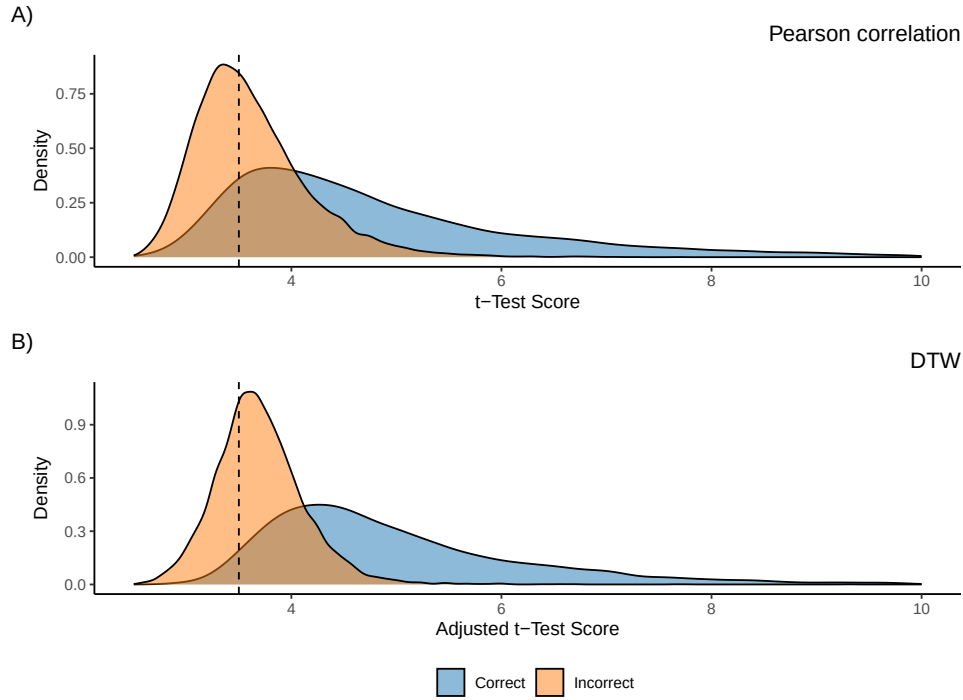


Figure 4.15: Densities of t -Test scores for correct (blue) and incorrect (yellow) alignments, for Pearson correlation (Panel A) and DTW (Panel B). The t -Test threshold of 3.5 is shown by a dashed line.

Examining the distribution of α scores in Figures 4.8 and 4.9 shows that very few values of α fall below 0.3 or exceed 0.8. This suggests some “compression” and implies that some adjustment to the calculation of α might be necessary. Also, the t -Test as calculated here does not account for insertions or deletions by DTW, which will increase α . Designing a new strategy to normalise DTW distances and compare alignments more robustly is another area for further experimentation.

4.4.4 Directions and Recommendations for Future Research

The results presented in this chapter demonstrate that DTW has clear potential in dendrochronology, particularly in cases of missing or added rings. For sample lengths above 50, typically employed as a minimum sample length, DTW performs as well as Pearson correlation, with significant improvements in cases where rings are added or removed. However, the algorithm and results presented are very much preliminary, and improvements and further research are necessary before implementation to “real world” dendrochronological problems. In light of the results of the

experiments presented in this chapter, the following recommendations are made:

- **Using different tree ring series for further testing of DTW**

DTW clearly outperforms Pearson correlation for longer samples from the Towai series, but before deploying DTW as a cross-dating method it is necessary to test the algorithm on other, perhaps *less* well-correlated, tree ring series.

- **Using an EPS threshold or similar in master sequence building and matching**

For similar experiments exploring alternative cross-dating methods, an EPS threshold applied during sampling from the tree ring series would offer more robust results. Using subsample signal strength (Buras, 2017) or simply series depth to constrain the sampling strategy might improve the accuracy of either alignment method.

- **Exploring and adjusting for biases**

Clear biases to certain regions of the Towai series are apparent in alignments for Pearson correlation (Figure 4.11). Exploring possible reasons for this bias, or whether similar effects are found for other tree ring series, is recommended. In addition, understanding why DTW is not so susceptible to this bias (Figure 4.12) might help to improve the accuracy of both methods.

- **Parameter tuning**

Further experiments exploring optimum values for k , z , and v parameters, or making them variable and application-specific (for example, varying v according to sample length), could improve the accuracy of DTW further.

- **Designing a method for finding missing or added rings**

Using non-linearities in the DTW warping path to detect missing or added rings would be reasonably straightforward, and would be a useful extension of the algorithm for dendrochronology.

- **Improved Accuracy Metrics**

This experiment confirms that the t -Test score and 3.5 threshold is not an effective accuracy metric for alignment by either DTW or Pearson correlation. Further exploration of alternative metrics for either method would certainly be worthwhile.

4.5 Conclusion

This chapter has shown that DTW has valuable potential for cross-dating in dendrochronology. By preparing thousands of samples for cross-dating and comparing between the methods, results suggest that DTW is an excellent alternative to Pearson correlation, particularly where missing or false rings are present or suspected in a sample. Alongside this contribution, this experiment highlights that the t -Test and its commonly applied threshold of 3.5 are inadequate as an accuracy metric. Correct and incorrect matches by either alignment method were not separated by t -Test scores, and further research into this problem is recommended. Although it is possible to find correct matches by either method for shorter sample lengths, both alignment methods were much more successful for sample lengths greater than or equal to 60. As such, the minimum sample length of around 50 rings typically required for cross-dating is confirmed as appropriate. Future research exploring the potential of DTW for cross-dating should focus on improving accuracy metrics, tuning DTW parameters, and exploring biases in alignment apparent by Pearson correlation.

5

Acid-Base-Acid with Stepped Combustion (ABA-SC) as a Pretreatment Method for Radiocarbon Dating of Charcoal

Abstract

This chapter presents results from an experiment in radiocarbon dating, exploring an alternative chemical pretreatment method for charcoal. At the Oxford Radiocarbon Accelerator Unit (ORAU), there are currently two routine laboratory pretreatment protocols for charcoal: Acid-Base-Acid (ABA); and Acid-Base-Oxidation with Stepped Combustion (ABOx-SC)¹. The latter is generally recognised as being more aggressive and in many contexts more effective at removing contamination from charcoal samples. The improved contaminant removal by ABOx-SC can be particularly significant when dating material older than 25000 ¹⁴C years. There are several fundamental differences between ABA and ABOx-SC protocols, and despite having both been routinely practised at the ORAU for over 15 years, it is unclear exactly *why* ABOx-SC is so effective, nor which aspect of the ABOx-SC protocol is most significant for contaminant removal in comparison to ABA.

¹An adapted ABOx-SC protocol omitting the base step (AOx-SC) is also routinely applied at the ORAU. See Chapter 6 for a description of the AOx-SC method.

In this study, the effect of the stepped combustion stage of ABOx-SC is tested and quantified using an adapted ABA protocol with stepped combustion (referred to at the ORAU as precombustion). Excess material from samples already pretreated by ABA at the ORAU was precombusted as for ABOx-SC. Seven charcoal samples from Grotta di Fumane in north-eastern Italy were used to test the ABA-SC method, having previously been shown to produce significant differences between their ABA and ABOx-SC dates. Excess ABA material for each sample was also re-dated without precombustion, providing an opportunity to compare duplicate ABA dates and assess the effect of up to 25 years of sample storage after pretreatment at the ORAU.

Radiocarbon determinations suggest that ABA-SC is unable to remove contamination as effectively as ABOx-SC. However, ABA-SC consistently produces older dates than ABA, and in five out of the seven samples from Fumane this effect was statistically significant. On average, and expressed as a proportion of the improvement in contaminant removal by ABOx-SC, precombusting ABA material increased contaminant removal by $46.5 \pm 23\%$. Indirectly, this implies that precombustion is a significant step in the ABOx-SC protocol. It is also possible that interaction between the precombustion and oxidation steps of ABOx-SC is another important aspect of contaminant removal in ABOx-SC. This is investigated in a separate experiment in Chapter 6. The results and implications of both experiments, as well as recommendations for future research, are discussed in Chapter 7.

5.1 Introduction

Charcoal is a widely used material for dating by radiocarbon. This is due to its high carbon content, its abundance in many archaeological and palaeoenvironmental contexts, and its resistance to post-depositional alteration in comparison to other routinely dated materials (Bird, 2013; Bird & Ascough, 2012). Charcoal is the carbon-rich product of the partial combustion of plant matter, formed by thermal decomposition under conditions of restricted oxygen, or pyrolysis. Under pyrolysis, the hydrogen, nitrogen, oxygen, and sulphur contents of plant material are reduced, leaving a highly stable carbon fraction with an aromatic molecular structure (Ascough et al., 2008, 2009; Ascough, Bird, Francis, Thornton, et al., 2011; Bird et al., 2014). A long-standing relationship between humans and fire (Roebroeks & Vill, 2011), along with an abundance of “natural” charcoal produced by biomass burning during wildfires (Zackrisson et al., 1996), mean that charcoal is frequently recovered at archaeological sites and is present in many palaeoenvironmental records (Bird, 2013). Given that charcoal is plant-derived, any remaining ^{14}C within a sample is biospheric in origin and therefore is not subject to reservoir effects, as ^{14}C measured in plants is largely in equilibrium with that of the atmospheric carbon system (Bird & Ascough, 2012; Hajdas et al., 2021). These factors in combination mean that charcoal has long-since been regarded as an excellent material for radiocarbon dating.

In radiocarbon dating, *endogenous* carbon is generally the fraction targeted for dating. For charcoal, the endogenous fraction is absorbed by plants from the atmosphere via photosynthesis, and begins to decay when a plant dies. *Exogenous* carbon, on the other hand, refers to carbon present in a sample that originates from any other source, and is typically considered to be contamination that should ideally be removed before radiocarbon measurement. Although often considered “inert” due to its aromatic stability and resistance to degradation, recent research has recognised charcoal as a heterogeneous material that is not immune to oxidation, chemical alteration, or interaction with the burial environment post-deposition

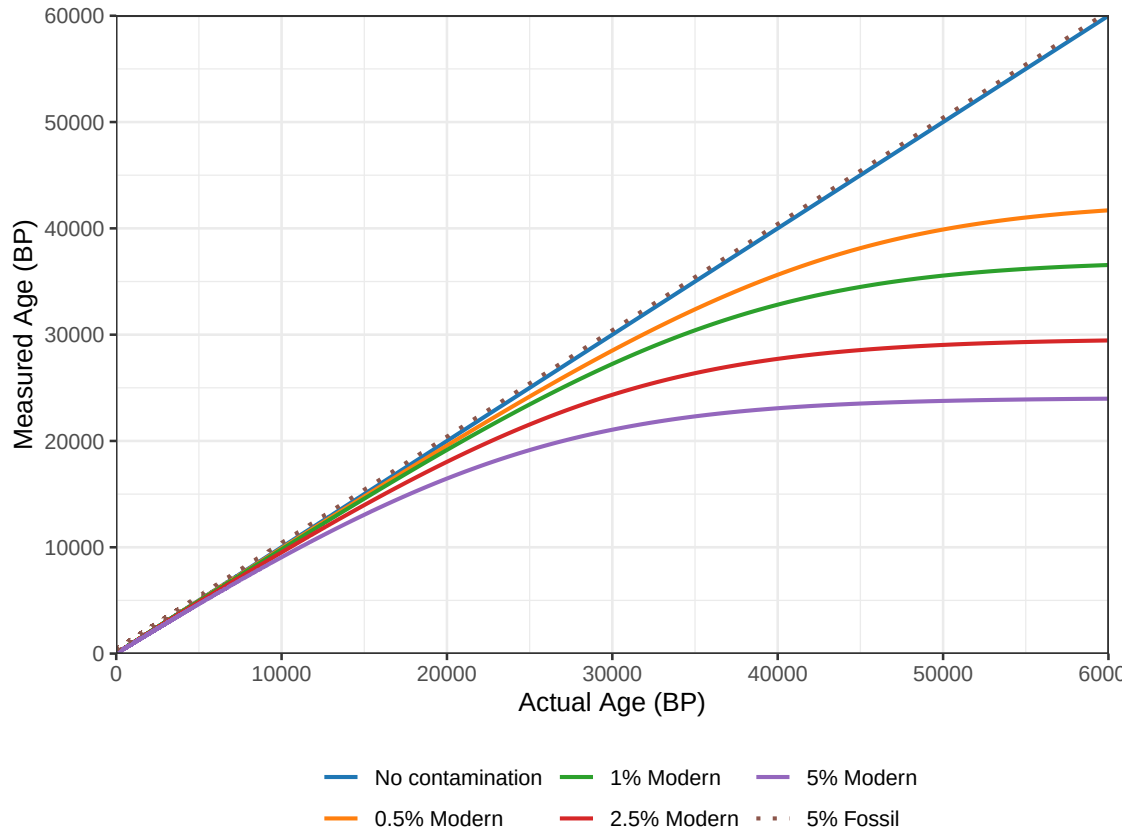


Figure 5.1: The effect of modern and fossil contamination in radiocarbon measurement. As the actual ^{14}C age of the sample increases, modern contamination has an increasingly significant effect on the measured ^{14}C age, whilst for the inclusion of fossil (^{14}C -dead", dotted line) contamination, the effect is linear.

(Ascough, Bird, Francis, Thornton, et al., 2011; Bird et al., 2008, 2017). An additional concern for radiocarbon dating is the material's high porosity and large external and internal surface areas, which can contribute its capacity to absorb exogenous carbon that is readily transported and cycled through and across the sediment column. These processes can deposit contaminant carbon to a charcoal sample from a range of exogenous sources, which in turn can have a dramatic effect on ^{14}C measurements, as shown in Figure 5.1 (Bird et al., 1999, 2014; Douka et al., 2010; Higham, Brock, et al., 2009; Higham, Barton, et al., 2009; R. Wood et al., 2012). The absorption of exogenous carbon by charcoal can be a significant challenge in obtaining accurate radiocarbon dates, and effective strategies towards its removal are of primary concern prior to radiocarbon measurement.

Generally, the presence of exogenous carbon within a sample poses more of

a problem for older samples than for younger ones. Contaminants deposited to charcoal post-deposition will in most cases be younger than the true age of the sample, and as for all materials dated by radiocarbon, the effect of younger and/or modern carbon included in a sample for dating becomes more significant as the true age of the sample increases (Pettitt et al., 2003; R. Wood, 2015). This issue can create a complex barrier to creating and understanding reliable radiocarbon chronologies, for Palaeolithic archaeology in particular (Higham, 2011; Higham et al., 2014). Charcoal produced close to the radiocarbon boundary (around 50000 ^{14}C years before present; ky BP)² that contains very small amounts of contaminant modern carbon can produce dating inaccuracies of thousands of years (Higham, Barton, et al., 2009; R. Wood et al., 2012). Although contamination by “old” or “fossil” carbon is a possibility, younger contamination has a much more significant effect on radiocarbon dates than the introduction of contaminant “ ^{14}C -dead” carbon, and the former is generally of much greater concern (R. Wood, 2015). These effects are summarised in Figure 5.1.

Chemical pretreatment protocols were designed very soon after the development of the radiocarbon dating method, in order to remove as much of the exogenous fraction from a sample as possible before radiocarbon measurement. Today, a variety of protocols are routinely applied to a broad range of materials, while their design and development continue to be very active areas of research (Ascough, Bird, et al., 2010; Bird et al., 2010; Brock et al., 2010; Gillespie, 2019; R. Wood, 2015). The dominant pretreatment methods designed to remove contamination in radiocarbon dating of charcoal are described in Section 5.1.2. Before this, some description and definition of the sources and causes of contamination in charcoal are provided.

²In this chapter, all ages “BP” or “ky BP” refer to conventional radiocarbon ages before present (1950). Ages on different timescales (e.g., calendar years, calibrated radiocarbon years, GICC05, etc.) are referred to and abbreviated separately.

5.1.1 Contamination in Charcoal

The origins and level of contamination in fossil charcoal can vary significantly. In order to understand the design and effectiveness of various pretreatment strategies intended to remove the exogenous fraction, it is important to first consider the possible origins of exogenous carbon, as well as the broader variety of contaminant types and factors that affect charcoal’s susceptibility to adsorb them.

For charcoal, there are a number of key exogenous carbon sources that are typical to archaeological and palaeoenvironmental contexts. A major source and vehicle of exogenous carbon within the sediment column is groundwater movement (Alon et al., 2002; Bird et al., 2002; Boaretto et al., 2008; Hajdas et al., 2021; Mellars, 2006). As the groundwater table fluctuates, water-soluble contaminants - often referred to as “humics”³ - can be transported to and adsorbed by charcoal surfaces. These contaminants can include a range of organic compounds, such as carbohydrates, amino acids, or phenolics, that can all be deposited to a charcoal sample after deposition (Bird, 2013; Pietikäinen et al., 2000). Organic compounds can also encourage microbial colonisation and activity, further increasing the possibility of recalcitrant carbon being deposited to charcoal samples (Ascough, Sturrock, et al., 2010; Bird et al., 2017; Tilston et al., 2016). In addition, surrounding sediment may also contain carbonates that can adhere to charcoal (Alex et al.,

³It is worth noting that within the radiocarbon community, common definitions of the various sources of contamination are somewhat complex, and at times vague. The case of “humics” is a good example. Following Head (1987), the humic fraction targeted for removal is defined at the ORAU as “the fraction extracted by alkaline solution that becomes insoluble after acidification” (Brock et al., 2010: 104). The “fulvic” fraction, meanwhile, is soluble in both acid and alkaline solutions, and the term “humins” describes the residue *insoluble* in both acid and alkaline solutions. The humin fraction is generally targeted for ¹⁴C measurement. A semantic issue here is that the “humic” fraction is defined by the product of extraction, rather than by the source(s) of contamination within the burial environment. Since the pretreatment protocol is non-targeted, the concept is circular: the “humic” fraction will *always* be removed from a charcoal sample after washing with acidic and alkaline solutions, because its very definition is that which is removed by acidic and alkaline washes. Lehmann & Kleber (2015) note that, for sediments, the process of humic extraction is neither exact nor targeted. Biomass, dissolved organic matter, or undecomposed plant material are all potentially extracted by an alkaline solution, in varying proportions. Similarly, “fulvic acids” are a non-specific category without clear definition. Therefore, with regards to contaminant extraction in radiocarbon dating, more appropriate labels for categories referred to as “fulvic acids” and “humic acids” are “acid-soluble” and “alkaline-soluble” organic fractions, respectively.

2017; Rebollo et al., 2011). This broad array of contaminant sources all have potential to transport exogenous carbon to a charcoal sample, that in turn have an effect on radiocarbon measurements.

A number of studies have demonstrated that the conditions of pyrolysis can have significant effect on the susceptibility of charcoal to chemically interact with the burial environment. In relation to this, Bird & Ascough (2012) describe a “combustion continuum” for pyrolysed material, referring to the varying conditions of pyrolysis and density of material burned in charcoal production, which in turn have an effect on its chemical stability after deposition (Ascough et al., 2008; Bird et al., 2017; M. W. I. Schmidt & Noack, 2000). Pyrolysed plant material produced at lower temperatures ($< 350\text{ }^{\circ}\text{C}$; e.g., slightly charred plant remains) has low stability, low aromaticity, and high reactivity, and therefore is more likely to be lost or degraded in the burial environment after deposition, as well as being more susceptible to exogenous contamination. Pyrolysed material formed at higher temperatures ($> 400\text{ }^{\circ}\text{C}$) and by a greater density material, meanwhile, is more stable and resistant to environmental interaction post-deposition (Bird et al., 2017; Bird & Ascough, 2012; Braadbaart et al., 2009; Dufourny et al., 2019; Tilston et al., 2016). Charcoal’s long-term stability and its capacity to interact with the burial environment are of primary concern for radiocarbon dating, but these characteristics are rarely known prior to radiocarbon measurement.

Oxidation and the duration of exposure to the burial environment are additional factors that influence charcoal contamination and preservation (Bird & Ascough, 2012). Comparisons between Fourier transform infrared spectroscopy (FTIR) spectra of modern and fossil charcoal show that carboxylate groups identifiable in fossil charcoal are not found in modern samples, suggesting that oxidation is a key aspect of charcoal degradation over long periods (Boaretto et al., 2008; Cohen-Ofri et al., 2006, 2007). This process is referred to as oxidative weathering, and involves the addition of oxygen to the aromatic skeleton of charcoal, also inferred by relatively high O/C and O/H ratios (Ascough, Bird, Francis, Thornton, et al., 2011, 2011; Cohen-Ofri et al., 2006; Tomiyama et al., 2016). Material that

has undergone significant oxidation is more likely to be poorly preserved and less likely to survive chemical pretreatment, particularly during dramatic changes in pH between acid and base washes.

On Bird & Ascough (2012)’s combustion continuum, significantly oxidised charcoal is likely to be produced by a low-density plant material or at low temperatures, but is also more reactive, and therefore is also more likely to suffer from contamination. By contrast, if the charcoal is well-preserved and is not significantly degraded by oxidation, its structure will have more hydrophobic surfaces due to a higher proportion of graphite-like crystallites, and is therefore less likely to adsorb or be affected by hydrophilic contaminants transported by groundwater (Cohen-Ofri et al., 2006; Tomiyama et al., 2016). As such, the problem of pretreatment (see Section 5.1.2) becomes a question of balance: designing protocols that are aggressive enough to remove a broad range of contaminants from a variety of charcoal types, but not so destructive that poorly-preserved samples are destroyed.

Chemical bonds between exogenous and endogenous material are likely to occur at the edges of the charcoal’s molecular structure (Ascough, Bird, Francis, Thornton, et al., 2011; Tomiyama et al., 2016). If loosely bound, this exogenous material is readily removed by base washes that target the exogenous fraction in chemical pretreatment (see Section 5.1.2). However, if the chemical bonds created during contaminant adsorption are strong, the exogenous fraction will be more difficult to remove and may remain tightly bound after pretreatment. Again, the process of post-depositional alteration and the likelihood of chemical bonds being strong appear to be - at least in part - dependent on the degree of oxidative weathering and the conditions of pyrolysis (Ascough, Bird, Francis, Thornton, et al., 2011).

A complication here is that as charcoal undergoes post-depositional oxidation, its structure increasingly resembles that of alkaline-soluble contaminants. This process is termed by Cohen-Ofri et al. (2006) as “self-humification.” As a result, alkaline washes in chemical pretreatments designed to remove the exogenous (i.e., “humic”) fraction from a sample may in fact remove large amounts or all of the charcoal material, along with the endogenous carbon fraction targeted for dating.

Ascough, Bird, Francis, & Lebl (2011) confirm, for example, that often one of the dominant components of material extracted during chemical pretreatment appears to be the charcoal itself. This highlights the possibility that a large proportion of the alkaline-soluble fraction is actually an endogenous product, rather than any contaminant. Removing endogenous material along with contaminants is not a problem if long ample proportions can be retained for dating, but dating very poorly-preserved charcoal can be challenging, as it is very difficult to know whether dates reliably relate to the endogenous fraction, or samples may be lost altogether during pretreatment chemistry.

The conditions and materials of charcoal production, along with the conditions of deposition and post-depositional groundwater and soil microbial activity, as well as interaction between these factors, can all have considerable effect on the proportions of endogenous and exogenous carbon present within a charcoal sample (Bird et al., 2014, 2017). In light of this variability, it follows that chemical pretreatment methods should also require some variability and sample-specific considerations. In the following section, the predominant pretreatments for charcoal are described.

5.1.2 Pretreatment Methods for Radiocarbon Dating of Charcoal

Since the 1950s, routine chemical pretreatments designed to remove the exogenous fraction from charcoal have involved a combination of acid and base washes (De Vries & Barendsen, 1954). The most common protocol is known as “ABA” (Acid-Base-Acid)⁴, and is the charcoal pretreatment method most often implemented at the ORAU (see Section 5.2.2 for a full description of the ORAU’s ABA protocol). Considering the variability of contaminant sources described in Section 5.1.1, the ABA pretreatment is logical in its design: an acid wash removes sediment-derived carbonates not removed during initial sampling; a base wash targets organic contaminants deposited to the sample through groundwater table movement and microbial activity; then a final acid removes any atmospheric CO₂ absorbed by

⁴The ABA method is sometimes referred to as “AAA” (Acid-Alkaline-Acid).

samples during the base wash (Bird, 2013; Brock et al., 2010; De Vries & Barendsen, 1954; Hatté et al., 2001).

Exact ABA protocols can differ significantly between laboratories, such as the duration and strength of acid and base washes, their temperatures, or whether base washes are repeated (Bird et al., 1999; Brock et al., 2010; Hajdas et al., 2004; Kim et al., 2001; Mayer et al., 2008; Sakamoto et al., 2004). However, Santos & Ormsby (2013) observe little difference in radiocarbon determinations when comparing five distinct versions of the ABA protocol, suggesting that these discrepancies have little effect in practice. Following its theoretical logic, the ABA method *should* be sufficient to remove contaminants from charcoal.

However, numerous studies have indicated that, in certain contexts, exogenous contamination can remain problematic when radiocarbon dating charcoal samples by ABA (Bird et al., 1999; Brock & Higham, 2009; Douka et al., 2010; Gillespie, 1997; Higham, Barton, et al., 2009; Higham, Brock, et al., 2009; R. Wood et al., 2012). This has been particularly apparent for charcoal from contexts older than c. 25 ky BP and samples deposited in highly contaminant environments. Where samples are suspected to be older than 25 ky BP or highly contaminated, the routine ABA pretreatment method is sometimes replaced with a more rigorous chemical process known as “ABOx-SC” (Acid-Base-Oxidation with Stepped Combustion), originally designed by Bird et al. (1999) and shown to remove a higher proportion of exogenous carbon than ABA.

The ABOx-SC protocol as implemented at the ORAU involves much stronger acid and base washes than ABA, before an aggressive oxidation step using acidified potassium dichromate. The material remaining after these more aggressive treatments was described by Bird & Gröcke (1997) as “oxidation resistant elemental carbon,” or OREC. Following oxidation, the OREC samples were then subjected to “stepped combustion” by Bird et al. (1999). This involved sequential combustion at increasing temperatures (350, 630, and 850 °C), designed to explore the proportion of contamination in each fraction. Bird et al. (1999) found that fractions of CO₂ produced at lower combustion temperatures (350 and 630 °C) had higher

^{14}C contents than the high temperature fraction (850 °C). This suggests that labile exogenous carbon is preferentially combusted and removed during the low temperature combustion steps, and the high-temperature fraction contains a higher proportion of endogenous carbon. As such, stepped combustion is routinely - but not always - applied following the ABOx wet chemistry. The effect of the ABOx-SC method in some cases has been substantial, producing dates significantly older than routine ABA pretreatments in many Palaeolithic and Pleistocene contexts (Bird et al., 2014; Brock & Higham, 2009; Haesaerts, 2013; Higham, Barton, et al., 2009; R. Wood et al., 2012), although this is not always the case (Alex et al., 2017; Rebollo et al., 2011).

While ABA and ABOx-SC are currently the dominant pretreatment methods for charcoal and those routinely implemented at the ORAU, a number of alternative pretreatment methods are also of note. Hydrogen pyrolysis (“hypy”) is a relatively new technique that aims to separate the pyrogenic component of charcoal from almost all labile organic matter, by pyrolysis at ~ 600 °C using high-pressure hydrogen gas (Ascough et al., 2009; Ascough, Bird, et al., 2010; Bird et al., 2014). Comparisons to ABA and ABOx-SC show that contamination removal by hypy alone is not sufficient for samples close to the radiocarbon limit (Bird et al., 2014), although combining the method with ABA (ABA/hypy) has recently shown promising results, as well as developments in rapid throughput methods (Haig et al., 2020; Orr et al., 2021). Plasma ashing has also been experimentally applied to samples, and could be useful where preservation and significant sample loss is of a concern, due to its reliability and targeted nature (Bird et al., 2010). However, neither of these methods has been able to replicate dates as old as ABOx-SC with any statistical certainty, suggesting their effectiveness for contaminant removal is inadequate for heavily contaminated samples.

Although obviously effective, it is not exactly clear *why* ABOx-SC produces older dates than ABA, hypy, or plasma ashing in some contexts. As described in Section 5.1.1, charcoal contamination can be caused by a complex interaction of chemical and biological factors, and can vary significantly between samples and

contexts. Without understanding how ABOx-SC works to target and remove such a range of contaminants so effectively, it is difficult to explain why ABA does not work in some contexts, nor why in some cases ABOx-SC is *not* the most effective pretreatment (Alex et al., 2017; Rebollo et al., 2011; R. Wood, 2015).

Unpacking this further, there are multiple methodological differences between ABOx-SC and ABA. Increased acid and base concentrations, the introduction of acidified potassium dichromate, the use of stepped combustion, or any number of these factors in combination could be the reason for substantial differences between dates produced by both methods. Currently, a comprehensive exploration and comparison of these factors, using adequate and appropriate material, is lacking.

It is possible that stepped combustion alone is the primary component of improved contaminant removal in ABOx-SC: in Bird et al. (1999)’s original study, stepped combustion had a significant effect on radiocarbon dates for untreated, ABA-, and ABOx-treated samples. The explanation given for this by Bird et al. (1999), as well as subsequent applications of the ABOx-SC technique, is that labile exogenous carbon is combusted to CO₂ more readily and preferentially at lower temperatures, as opposed to the more stable aromatic carbon structures endogenous to the sample that are formed during the original pyrolysis event (Bird et al., 2014; Tomiyama et al., 2016). Following this logic, ABA with stepped combustion (ABA-SC) should also show significant differences to dates produced by ABA, particularly for samples that have previously been shown to produce substantially older dates using ABOx-SC. If ABA-SC can remove contaminants to the same degree as ABOx-SC, laboratory processing times could be significantly reduced, as the laboratory logistics of using potassium dichromate in ABOx-SC are laborious and often impractical (see Chapter 6).

An alternative hypothesis is that chemical oxidation by potassium dichromate *in combination with* stepped combustion is responsible for the more effective removal of contaminants by ABOx-SC. Exogenous carbon may only become more labile and readily combusted *after* oxidation in acidified potassium dichromate, and therefore ABA-SC would not remove any more contamination than ABA. In

other words, stepped combustion may only be relevant in combination with an aggressive oxidation step, in which case ABA-SC dates would be indistinguishable from their ABA counterparts.

A comprehensive comparison between radiocarbon measurements produced by ABA-SC and ABOx-SC may illuminate whether stepped combustion is the dominant factor in the ABOx-SC protocol. If ABA-SC can produce similar results to ABOx-SC, it is likely that stepped combustion alone has a significant effect on radiocarbon measurements, and the oxidation and aggressive acid and base washes of ABOx-SC are unnecessary. In the following section, these hypotheses are considered in more detail, through a review of previous studies comparing ABA-SC and ABOx-SC methods.

5.1.3 A Review of Previous Studies Comparing ABA-SC and ABOx-SC

A number of previous studies have applied ABA with stepped combustion in chemical pretreatment, at times providing direct (i.e., paired) comparison to dates produced by ABOx-SC, and in some cases further comparison to ABA. These studies are summarised in Table 5.1.

The original ABOx-SC study by Bird et al. (1999) included a comparison of ABA and ABOx dates, both with stepped combustion at 330 °C, 630°C, and 850 °C. Only one sample was dated by both ABA-SC and ABOx-SC, of “¹⁴C-dead” (i.e., from contexts beyond the radiocarbon limit, > 50 ky BP) charcoal from a buried palaeosol in Canberra, Australia. The sample was known to have been subject to groundwater fluctuation and contamination, and was overlain by a gravel unit, stratigraphically correlated to a gravel fill in an adjacent valley with an optically stimulated luminescence (OSL) age of greater than 50,000 years BP (kyr BP, Bird et al., 1999). Although Bird et al. (1999) argue that compared dates for this sample indicate ABA-SC is not sufficient to remove contaminants, ABA-SC and ABOx-SC dates for the 850 °C fractions are in fact statistically indistinguishable (ABA-SC: 52230 ± 1430 BP; ABOx-SC: 55860 ± 2350 BP; $\chi^2 = 3.125$, df = 1, N = 2, $p = <$

Table 5.1: Previous studies using ABOx-SC and ABA-SC pretreatment methods on paired samples. Column N relates to the number of samples dated by both ABOx-SC and ABA-SC.

Reference	Material Dated	Known Age	N	Context	Agreement	Notes
Bird et al. 1999	Charcoal	> 50 ky BP	1	Groundwater contaminated	Statistically indistinguishable	Untreated, stepped-combusted sample also produced background age
Turney et al. 2001	Lake sediment	0 - > 50 ky BP	10			
Santos et al. 2001	Wood	> 50 ky BP, c. 20 ky BP, c. 40 ky BP	5		All but one pair statistically indistinguishable	
Santos et al. 2013	Wood	> 50 ky BP	3		Statistically indistinguishable	
Tomiyama et al. 2016	Charcoal	Age not known	3		Statistically indistinguishable	
Alex et al. 2017	Charcoal	Age not known	4	Clay-rich	ABA produced earlier dates	
Bird et al. 2014	Charcoal	> 50 ky BP		Known contaminant		Stepped-combustion not used

0.1)⁵. Furthermore, both ¹⁴C ages are greater than 50 ky BP, generally considered to be the limit of reliable measurement for radiocarbon determinations. It is also worth noting that the 850 °C fraction for an *untreated* and stepped combusted sample of the same charcoal produced a radiocarbon age above 50 ky BP, at 50360 ± 1700 BP. For this particular sample, stepped combustion alone seems to have removed sufficient contamination to give a comparable age to dates produced by ABOx-SC. This suggests that either stepped combustion is a significant factor, or that the sample was not sufficiently contaminated to allow an adequate comparison between methods for contaminant removal.

Turney, Bird, et al. (2001) use both ABA-SC and ABOx-SC to date sediments from Lynch’s Crater, a lake sediment record from north-eastern Australia. Macroscopic charcoal is present and abundant through the upper part of the

⁵All radiocarbon measurement errors reported in this chapter are 1 σ .

Lynch's Crater sequence, but towards the base of the core charcoal is found only in fine grains. Comparisons between the two methods show general agreement for the upper part of the core (< 11 m), but chronological inconsistencies are evident in ABOx-SC dates for the lower part of the sequence (> 40 ky BP). The authors suggest this is probably due to variable amounts of microbial contamination towards the base of the core, coupled with problematic ABOx-SC samples as indicated by low percent carbon (%C) yields, as might be expected for bulk mixes of sediment and fine-grained charcoal. [Turney, Bird, et al. \(2001\)](#) note that ABOx-SC samples were treated by acidified potassium dichromate for variable lengths of time, which may have resulted in variable amounts of contaminant removal between samples. Overall, the authors consider the chronology produced by ABA-SC dates to be more robust, with greater consistency and lower background ^{14}C variation compared to ABOx-SC.

[Santos et al. \(2001\)](#) test the application of ABOx-SC, ABA-SC, and cellulose extraction with stepped combustion (CE-SC), on a number of wood samples known to be ^{14}C -dead, as well as a sample below the Rotoehu Ash eruption in New Zealand, previously dated by OSL to 42 - 44 ka yrs BP. All but one of the samples produced comparable dates for ABA-SC and ABOx-SC methods, the exception being ANU-918, a sample of kauri log from a "strongly humified" sand deposit in New South Wales, Australia. ANU-918 produced dates of 41580 ± 620 ky BP for ABA-SC, 48000 ± 1830 for ABOx-SC, and 52290 ± 1830 for CE-SC. In the case of ANU-918, ABOx-SC and CE-SC appear to have removed contaminants from wood samples more effectively than ABA-SC, but considering other samples in the study, the effect of stepped combustion in contaminant removal does not appear to be consistent.

In another set of experiments on fossil wood, [Santos & Ormsby \(2013\)](#) compared ABA-SC and ABOx-SC for two ^{14}C -dead samples. The majority of dates in the study were produced to explore variability in ABA protocols between laboratories, meaning that only three samples had both ABA-SC and ABOx-SC treatment: one from South Auckland, New Zealand, from beneath a flood deposit dated $0.49 \pm$

0.05 Ma; and two samples from Airedale Reef, New Zealand, from underneath avalanche debris inferred to have been deposited during Oxygen Isotope Stage 5c. All samples were from contexts well beyond the radiocarbon boundary, and all pairs of ABA-SC and ABOx-SC dates produced indistinguishable results, beyond 50 ky BP. However, for all samples use in the study, ABA *without* stepped combustion also produced comparable ages to ABA-SC and ABOx-SC, suggesting that, for these samples, ABA was sufficient to remove contamination to the same degree as other pretreatment methods.

[Tomiyaama et al. \(2016\)](#) provide one of the most comprehensive and relevant chemical comparisons between ABA and ABOx methods. The authors produce dates for charcoal produced at various temperatures, using FTIR and carbon, hydrogen, and oxygen elemental analyses, all at multiple stages of the ABA and ABOx pretreatments. Their study included ABOx-SC and ABA-SC comparisons for three charcoal samples from a single charred branch, recovered from pyroclastic flow deposits in southern Kyushu, Japan. Although these were not known-age samples, contamination was inferred from the depositional context and further suggested by significant differences between radiocarbon dates the various pretreatment methods applied, as well as for untreated samples. All three pairs of dates for ABA-SC and ABOx-SC were statistically indistinguishable, with both methods consistently producing older dates than any other pretreatment without stepped combustion.

Although the sample size is limited, [Tomiyaama et al. \(2016\)](#)'s study suggests that stepped combustion has a substantial effect in contaminant removal. The authors concluded that oxidation in ABOx is more effective than the base step of ABA at removing exogenous organic matter bound to the exterior of aromatic structures of charcoal, by breaking the charcoal's molecular structure and isolating the inner aromatic components. Stepped combustion appears to have negated this difference, however, by removing high proportions of remaining contaminants after wet chemistry for either method. If this is the case, ABA-SC may be sufficient to remove exogenous contamination from most charcoal samples, and the oxidation steps in ABOx-SC may be unnecessary.

It is also worth noting that ABOx-SC does not always produce the oldest date. As noted in Section 5.1.1, the ABOx wet chemistry can remove substantial amounts of endogenous carbon from samples (Ascough, Bird, Francis, & Lebl, 2011; Cohen-Ofri et al., 2006). If samples are heavily contaminated with clay, this exogenous material can survive chemical pretreatment, while endogenous aromatic carbonic rings are broken down and the endogenous carbon removed by oxidation by ABOx. The result is a concentration of exogenous carbon, rather than endogenous. Alex et al. (2017) found that ABA consistently produced the oldest date compared to ABA-SC and ABOx-SC for four charcoals samples at Manot Cave, Israel, attributed to heavy clay contamination. FTIR analysis confirmed this effect, where the spectrum for an ABOx treated (without stepped combustion) sample showed more prominent peaks than ABA at wavelengths related to the presence of quartz and clay (Alex et al., 2017). Similar results were found by Rebollo et al. (2011) at Kebara Cave, Israel, where low percent carbon (%C) values for ABOx-SC dates suggested that material *other* than charcoal was concentrated during chemical pretreatment.

Other than the studies listed above, no other direct comparisons between ABA-SC and ABOx-SC have been made. While some of the previous studies comparing these methods use a great *depth* of techniques to examine each stage of chemical pretreatments (e.g., Tomiyama et al., 2016), their studies lack *breadth*, in that low numbers of samples were compared (just three in the case of Tomiyama et al., 2016). Given the complexity of the contamination process (Section 5.1.1), it is crucial that experiments in this topic use adequate numbers of samples, ideally from a variety of contexts. Earlier comparisons of ABOx-SC to ABA-SC that have paired a greater number of samples showed either that almost all results were indistinguishable (Santos et al., 2001), or that an ABA-SC chronology appears more robust (Turney, Bird, et al., 2001; Turney, Kershaw, et al., 2001). However, these studies compared dates on sediments or wood rather than charcoal, and these very different materials adsorb contaminants in distinct ways.

A comprehensive study comparing ABA, ABA-SC, and ABOx-SC, designed with appropriate charcoal samples and leading to a statistically robust conclusion

is currently lacking. Although ABOx-SC has produced older dates than ABA in many archaeological and environmental contexts, it is still unclear why this is the case ([R. Wood, 2015](#)). In many of the studies above ABA-SC appears to produce similar results to ABOx-SC. It is very possible that stepped combustion is the dominant factor for contaminant removal in ABOx-SC, and if ABA-SC can produce similar results, the aggressive and laborious wet chemistry and oxidation of ABOx-SC may be unnecessary.

5.1.4 Summary

The introduction to this chapter has considered:

1. The recognised sources of contamination (exogenous carbon) in charcoal and some problems with their definition and removal;
2. The main pretreatment methods for radiocarbon dating of charcoal (ABA and ABOx-SC);
3. Previous studies considering the effect of ABA with stepped combustion (ABA-SC) in comparison to ABOx-SC.

To reiterate, [Bird et al. \(1999\)](#)'s ABOx-SC introduced three key differences in comparison to the routine ABA pretreatment method:

1. Increasing the concentrations of the acid and base solutions;
2. Replacing the second acid wash with an oxidation step, using acidified potassium dichromate;
3. Using stepped combustion before graphitisation.

Although for many contexts ABOx-SC is accepted to be a more reliable pretreatment method for older charcoal, particularly where “humics” and contamination are likely to cause issue for radiocarbon dating, it is unclear which of the three differences described above has most importance in producing older radiocarbon dates than those by ABA.

Stepped combustion appears to have a significant effect on charcoal pretreated by a range of chemistry protocols. This chapter presents an experiment in charcoal pretreatment that aims to investigate and quantify the effect of precombustion for contaminant removal, by re-dating material previously pretreated by ABA at the ORAU with an added precombustion step. The results provide further comparisons between samples pretreated by multiple pretreatment methods: ABA; ABA-SC; and ABOx-SC. In particular, the following questions are considered:

1. *How much exogenous carbon does stepped combustion remove during chemical pretreatment?*
2. *Can ABA with stepped combustion (ABA-SC) produce dates similar to ABOx-SC?*

Table 5.2: Samples used in the experiment and their previous ABA and ABOx-SC radiocarbon determinations. ABA dates were made between 1996 and 2008 (column Year), while ABOx-SC dates were made 2007/8. χ^2 values for paired dates are given in the final column (df = 1, N = 2, p < 0.05) .

Sample	Context	ABA					ABOx-SC					χ^2
		Year	$\delta^{13}\text{C}$	%C	Result	$\pm$	$\delta^{13}\text{C}$	%C	Result	$\pm$		
Fumane												
9395.0	D3ba	1998	-24.8	62.2	32020	340	-24.8	68.7	33890	220	21.32	
22795.0	A2	2008	-25.6	60.7	32530	240	-24.2	61.8	34940	280	42.71	
13021.0	A2	2002	-23.3	44.3	31830	260	-21.7	74.5	35180	220	96.75	
13020.0	A2	2002	-25.2	58.5	30650	260	-22.5	80.6	35640	220	214.66	
22798.0	A2	2008	-24.7	64.4	33380	210	-23.8	40.5	35850	310	43.52	
7806.0	A5	1996	-22.1	62.3	33700	600	-21.1	74.4	40150	350	86.22	
9400.0	A5/A6	1998	-23.8	42.1	38800	750	-24.4	62.1	40460	360	3.98	
Maninjau												
22401.5		2008	-27.9	81.2	49400	2300	-24.4	88.7	53400	1500	2.12	
22401.6		2008	-25.2	82.4	50500	1000	-24.7	88.2	50600	1000	0.00	

5.2 Materials and Methods

5.2.1 Samples

Samples used in the experiment are listed in Table 5.2.

5.2.1.1 Grotta di Fumane

All archaeological samples were from Middle and Upper Palaeolithic layers at Grotta di Fumane (Fumane), a key Palaeolithic site for understanding the first dispersal of anatomically modern humans (AMH) into Europe (Benazzi et al., 2014, 2015; Falcucci & Peresani, 2019).

Fumane is a cave site located in the Veneto region of north-eastern Italy at 350 m above sea level, comprising several cavities and tunnels within a dolomitic limestone complex that have yielded over 10 m of sedimentary deposits (Peresani et al., 2008). Archaeological material is distributed across the stratigraphy with varying intensity, with significant amounts of material recovered from Middle Palaeolithic,

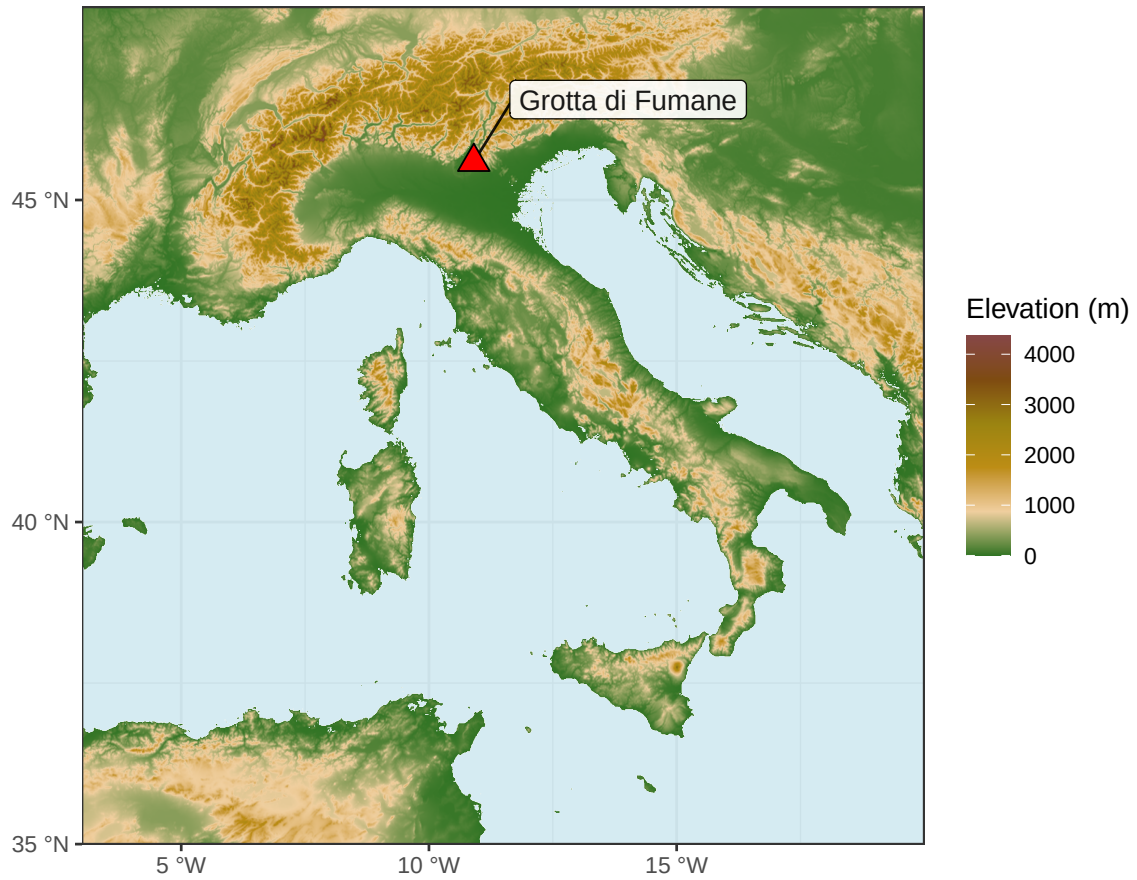


Figure 5.2: Location of Grotta di Fumane (red triangle).

Aurignacian, and Gravettian layers. The Middle to Upper Palaeolithic sequence has been divided into four main units: basal sands (B); breccia (BR); layers with considerable archaeological material (A); and the uppermost “slope-waste deposits and landslides” (D, Figure 5.3, Broglio et al., 2004: 263). Samples for this study were from final Mousterian (A5, A6), Aurignacian (A2), and Gravettian (D3ba) layers (Table 5.2; Broglio et al., 2004). For in-depth descriptions of the site stratigraphy and geological context see Broglio et al. (2004), Peresani et al. (2008), and Higham, Brock, et al. (2009).

As yet, no evidence of the Campanian Ignimbrite (CI) tephra horizon has been found at Fumane, precluding the use of a known-age horizon close to samples to verify the accuracy of methods as in other studies at Upper Palaeolithic sites such as Kostenki, Castelvita, and Serino (Douka et al., 2010; R. Wood et al., 2012). This means that the true or minimum ages of the samples were not known,

and the absolute accuracy of ABOx-SC or ABA-SC cannot be directly assessed in this context. However, relative accuracy can be inferred, where older dates are more likely to be closer to the true age of the sample (Figure 5.1; R. Wood, 2015). As mentioned above, this is because exogenous carbon is much more likely to be younger than the sample, therefore a method that produces older ages can be assumed to be more effective at removing the exogenous carbon fraction. Furthermore, over the past two decades an extensive dating campaign at Fumane has shown that ABOx-SC dates are in good agreement with dates obtained on bone by ultrafiltration, and together these measurements produced a Bayesian model in much closer agreement to the site's stratigraphy than a model using ABA dates (Higham, Brock, et al., 2009; Marín-Arroyo et al., 2023). As such, the existing ABOx-SC dates were assumed to be close to the true age of the samples from Fumane, and employed as a benchmark by which to compare experimental methods.

The archaeological samples from Fumane selected for this experiment have previously been shown to produce significant differences between dates produced by ABA and ABOx-SC ($\chi^2 > 3.84$, $df = 1$, $N = 2$, $p < 0.05$, Brock & Higham, 2009; Higham, Brock, et al., 2009). After pretreatment, yields can vary significantly, but excess material following wet chemistry is routinely archived at the ORAU. To test the effect of ABA-SC at Fumane, samples with sufficient excess material from previous ABA pretreatments ($> \sim 20$ mg) were precombusted and dated.

5.2.1.2 Maninjau Charcoal

In addition to the archaeological samples from Fumane, two ^{14}C -dead charcoal samples (standards) were included in the experiment, to assess whether modern carbon was added during precombustion. These samples were also previously dated by ABA at ORAU at the same time as some of the Fumane samples, and similarly had excess material after pretreatment chemistry that was archived in storage (Table 5.2). The charcoal standards were sampled from charred trees found within a tephra layer in the Maninjau Caldera, Sumatra (Alloway et al., 2004). Fission

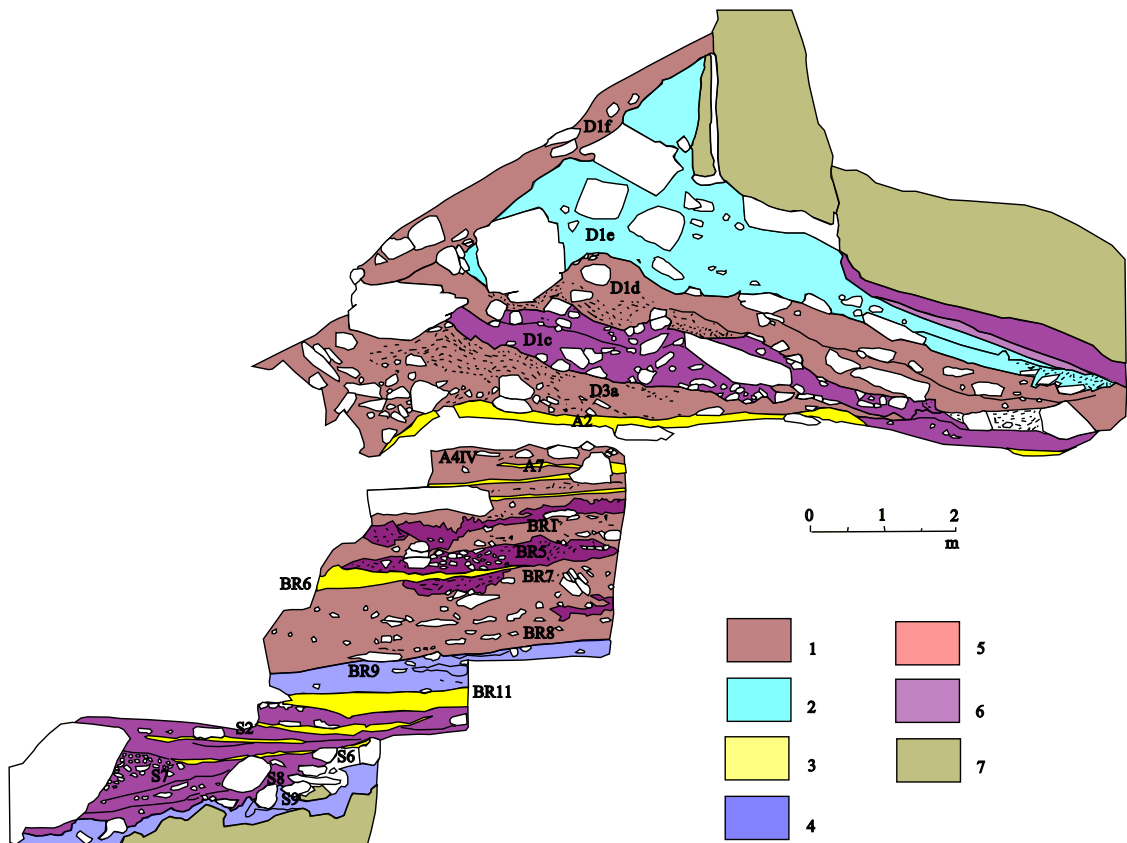


Figure 5.3: Stratigraphic diagram of relevant layers at Grotta di Fumane, Italy. Original image from Higham et al. (2009). Key: 1: denotes rendzina, upper soil; 2 slope deposits with boulders; 3 living floors, with high concentrations of organic matter or charcoal; 4 loess and sandy loess; 5 CaCO_3 cemented layers; 5 sandy deposits; 6 unweathered and weathered bedrock.

Track dating of glass shards within the tephra deposit date the Maninjav charcoal to 50 ± 3 ka BP. Radiocarbon dates *without* pretreatment are in agreement with these dates, suggesting the untreated charcoal is exceptionally clean and free of exogenous contaminants. As such, the Maninjav charcoal is an excellent background standard for assessing how much modern carbon is added during pretreatment (Alloway et al., 2004; Ascough et al., 2009; R. Wood et al., 2012).

5.2.2 ABA

All samples in Table 5.2 had previously been dated by ABA at the ORAU and had sufficient excess ABA material archived in storage for this experiment. The ABA pretreatment method currently applied at ORAU is described here for clarity, and

is assumed to be consistent for all samples since 1996 (Brock et al., 2010).

Charcoal fragments from bulk samples are hand selected, ideally as single entities and as free from environmental contamination and sediment as possible. Samples are roughly crushed before being washed in 1 M HCl at 80 °C for 20 minutes, followed by a base wash in 0.2 M NaOH at 80 °C for 20 minutes, before another acid wash at 1 M HCl at 80 °C for 1 hour. Between acid and base washes and after the final acid wash, the sample is rinsed 3 times with ultra pure MilliQ deionized water and filtered using Ezee-filters. Samples are then frozen at -18 °C, before freeze drying for a minimum of 12 hours. Following freeze-drying, samples for dating by ABA only are then combusted and graphitised in the routine manner for samples at ORAU (see below, Section 5.2.4).

5.2.3 Precombustion

There are some differences between the original stepped combustion procedure described by Bird et al. (1999), and the one currently routinely implemented at the ORAU (Brock et al., 2010; R. Wood & Brock, 2011). For clarity, *stepped combustion* here refers to the method described by Bird et al. (1999), and *precombustion* refers to the protocol implemented at ORAU (R. Wood & Brock, 2011). In principle, and in terms of the final AMS measurements obtained, the two methods are equivalent.

The current precombustion method at the ORAU is slightly different to the one described by Brock et al. (2010). Quartz wool separated into plugs is first heated under vacuum at 850 °C for 8 hours. Then, around 20 mg of chemically pretreated and freeze-dried material is transferred to a crimped quartz glass combustion tube, which is plugged by the quartz wool. The quartz glass tubes are evacuated, filled with high purity oxygen at 600 mbar, and sealed, before heating to 630 °C for 2 hours. Around 3 mg of the material product of this process is then graphitised by the routine method at ORAU (Brock et al., 2010).

This precombustion protocol differs to the stepped combustion originally described by Bird et al. (1999), where samples are sequentially combusted in the presence of copper oxide (CuO). In stepped combustion, samples are combusted

first at 330 °C for 2 hrs by O₂, then 630 °C for 1 hr, and twice at 850 °C for 1 hr. The CuO provides oxygen for higher temperature combustions. Fractions of CO₂ are captured at each combustion step, which in Bird et al. (1999)'s study were all graphitised for AMS dating. The final 850 °C step is repeated twice to ensure complete combustion and that all remaining carbon is collected for AMS measurement. Bird et al. (1999) showed that CO₂ captured from samples heated at 350 °C likely contains exogenous contaminants, and in subsequent applications of the ABOx-SC method this step has been skipped or the fraction discarded (Santos et al., 2001; Santos & Ormsby, 2013). CO₂ captured during the second combustion at 630 °C produced older dates than the 350 °C fraction, but the final combustion step's CO₂ produced consistently older dates than lower temperature combustions, and its product is thought to contain the highest fraction of carbon endogenous to the sample.

The ORAU *precombustion* protocol simplifies the stepped combustion as described by Bird et al. (1999), selecting and graphitising the endogenous carbon fraction that remains after combustion at 630 °C: the first combustion step is skipped, given that the CO₂ from this fraction is commonly discarded; the second combustion step is performed to *remove* and *discard* the 630 °C CO₂ fraction, rather than capture it for AMS measurement; and the third combustion step is indirectly performed by combusting the remaining product in the routine manner for samples at ORAU, at 900 °C.

Another key difference between Bird et al. (1999)'s method and that which is implemented at the ORAU is that, instead of using CuO to produce oxygen for oxidation during precombustion, the combustion tube is filled with 600 mbar of O₂. Internal experiments at ORAU have shown that, in terms of chemical composition (FTIR), precombusting with O₂ produces very similar products to those in the presence of CuO, with no significant difference in dates or $\delta^{13}\text{C}$ values (R. Wood & Brock, 2011). Furthermore, $\%C$ values are typically higher for precombustion using O₂, suggesting a more effective oxidation process than in the presence of CuO. As such, the precombustion protocol for the ABOx-SC samples dated in

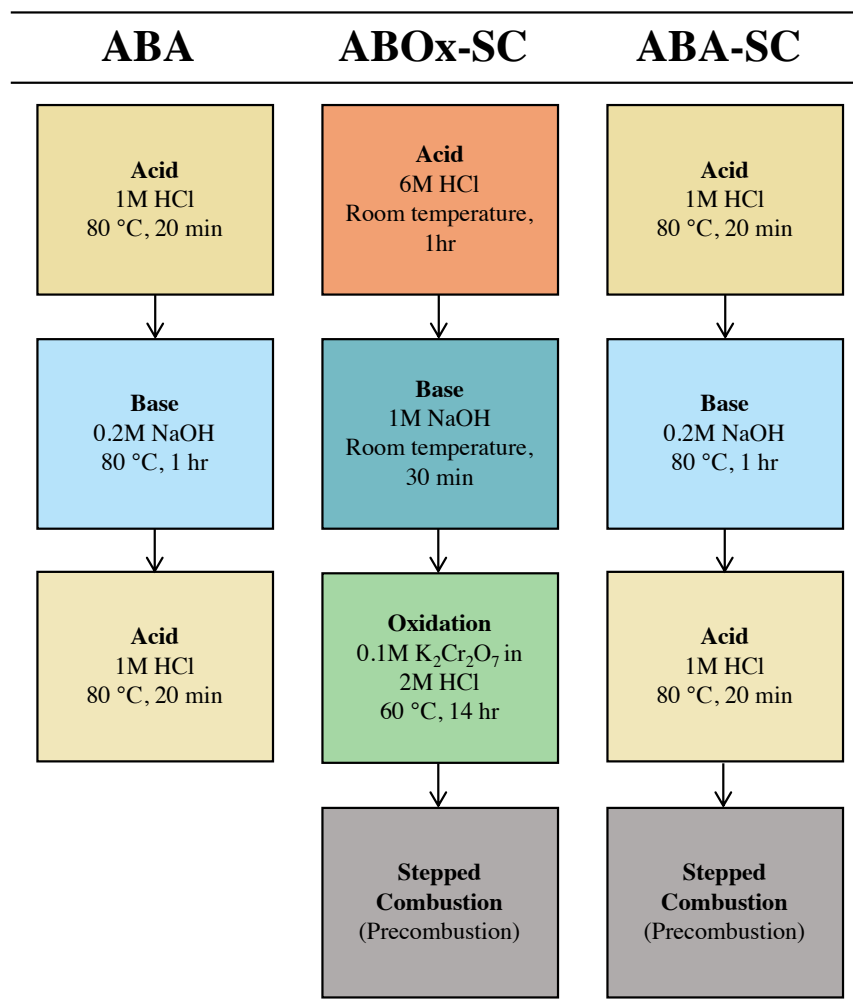


Figure 5.4: Schematic diagram of the ABA, ABOx-SC, and ABA-SC methods, as practiced at the ORAU.

2007/8 considered in this experiment differs to the precombustion protocol used for ABA-SC, and this should be kept in mind when making direct comparisons between the two methods.

A schematic diagram of the ABA, ABOx-SC, and ABA-SC methods is shown in Figure 5.4. A full description of the ABOx-SC method is given in Chapter 6.

5.2.4 Graphitisation and AMS Measurement

Graphitisation and AMS measurement followed methods described by Brock et al. (2010). Samples are weighed into tin capsules and combusted in a CHN elemental analyser (Carlo-Erba NA, 2000) coupled to a gas source isotope ratio

mass spectrometer (IRMS; Sercon 20/20). N₂ and CO₂ are separated by a gas chromatography (GC) column, carried by a helium carrier gas stream. A 50:1 splitter sends 2% of N₂ and CO₂ gases to the IRMS for analysis. The remaining CO₂ is collected cryogenically for graphitisation and dating.

CO₂ gas collected by combustion is transferred to rigs containing 2-2.5 mg of iron catalyst that have been out-gassed in the presence of 500 mbar H₂ at 450 °C for 1 hr. H₂ is added to the rig in the ratio of 2.2 H₂: CO₂, before heating at 560 °C for 6 hr to graphitise the sample. Graphitised material was pressed to aluminium targets before dating by AMS in the routine manner at the ORAU ([Brock et al., 2010](#)).

5.2.5 Calculating the Proportional Effect of Stepped Combustion in Comparison to ABOx-SC

In order to quantify the effect of precombustion and compare contaminant removal by ABA-SC between samples, the improvement in contaminant removal by ABA-SC over ABA was calculated, expressed as a proportion of contaminant removal by ABOx-SC. The proportional effect of precombustion (R) for each sample is given by

$$R = \frac{b - a}{c - a} \quad (5.1)$$

where a , b , and c are radiocarbon measurements for ABA, ABA-SC, and ABOx-SC, respectively, in ¹⁴C years. The uncertainty of R for a given sample, σ_R , is calculated using the derivatives of R with respect to a , b , and c :

$$\begin{aligned} \frac{dR}{da} &= -\frac{c - b}{(a - c)^2} \\ \frac{dR}{db} &= \frac{1}{c - a} \\ \frac{dR}{dc} &= -\frac{b - a}{(c - a)^2} \end{aligned}$$

σ_R is given by:

$$\sigma_R = \sqrt{\left(\frac{dR}{da}\sigma_a\right)^2 + \left(\frac{dR}{db}\sigma_b\right)^2 + \left(\frac{dR}{dc}\sigma_c\right)^2} \quad (5.2)$$

where σ_a , σ_b , and σ_c are the errors associated with radiocarbon measurements by ABA, ABA-SC, and ABOx-SC, respectively. To summarise the overall effect of precombustion, the mean and standard deviation of R across all samples was used.

5.2.6 Duplicate ABA Measurements

In some cases, samples had been archived for over 25 years at the ORAU. Since 1996, the earliest ABA measurement for samples in this study (Table 5.2), the ORAU has updated its AMS twice. To test whether ABA-treated samples had absorbed modern carbon during this time, and to avoid problems of discrepancies between instruments, updated AMS measurements were also made from the ABA excess (i.e., without precombustion). As such, a total of 18 new AMS measurements were made for the purposes of this study, 9 of which were repeated ABA dates.

Table 5.3: New radiocarbon measurements produced in this experiment for ABA (2023) and ABA-SC pretreatment methods, for archaeological samples from Fumane and Maninjau background standards, along with stable carbon isotope ($\delta^{13}\text{C}$) and percent carbon (%C) measurements. ABOx-SC measurements made in 2007/8 are also shown for comparison.

		ABOx-SC (2007/8)				ABA (2023)				ABA-SC (2023)			
Sample	Context	%C	$\delta^{13}\text{C}$	Result	$\pm$	%C	$\delta^{13}\text{C}$	Result	$\pm$	%C	$\delta^{13}\text{C}$	Result	$\pm$
Fumane													
9395.0	D3ba	68.74	-24.8	33840	310	62.41	-25.0	32560	250	92.79	-24.3	33500	210
22795.0	A2	61.75	-24.2	34880	400	64.01	-25.6	33110	290	84.00	-25.3	34550	340
13021.0	A2	74.46	-21.7	35180	220	62.75	-23.2	31350	220	89.39	-22.7	32400	200
13020.0	A2	80.59	-22.5	35640	220	57.77	-25.2	31240	210	85.18	-25.8	32140	200
22798.0	A2	40.51	-23.8	35750	600	62.87	-24.7	32850	320	78.74	-23.8	34040	590
7806.0	A5	74.41	-21.1	40050	550	61.55	-22.2	35110	320	93.84	-21.5	36430	330
9400.0	A5/A6	62.09	-24.4	40460	360	62.78	-24.2	38270	470	87.76	-24.1	38910	400
Maninjau													
22401.5						80.36	-27.8	49400	2300	86.63	-27.0	50100	1700
22401.6						79.83	-25.1	48700	1400	79.44	-23.5	49800	1500

5.3 Results

5.3.1 Radiocarbon Measurements

All new radiocarbon measurements made in this study (duplicate ABA dates and ABA-SC dates) are presented in Table 5.3 and Figure 5.5. Previous ABOx-SC measurements, made in 2007 and 2008, are also provided for comparison.

Maninjau standards showed background ^{14}C values beyond or within error of 50 ky BP for all new radiocarbon measurements. This suggested that no exogenous or contaminant radiocarbon was added to the samples before AMS measurement: during wet chemistry, precombustion, combustion, or graphitisation. χ^2 statistics for paired ABA and ABA-SC dates for all samples are shown in Table 5.4. Maninjau standards did not show statistically significant difference between ABA and ABA-SC dates, both with χ^2 scores less than 0.3 (df = 1, N = 2, p = < 0.05).

^{14}C dates for archaeological charcoal pretreated by ABA-SC showed a consistent pattern in comparison to ABA and ABOx-SC. For all samples from Fumane, ABOx-SC produced the oldest dates, followed by ABA-SC, then ABA (Figure 5.5). For

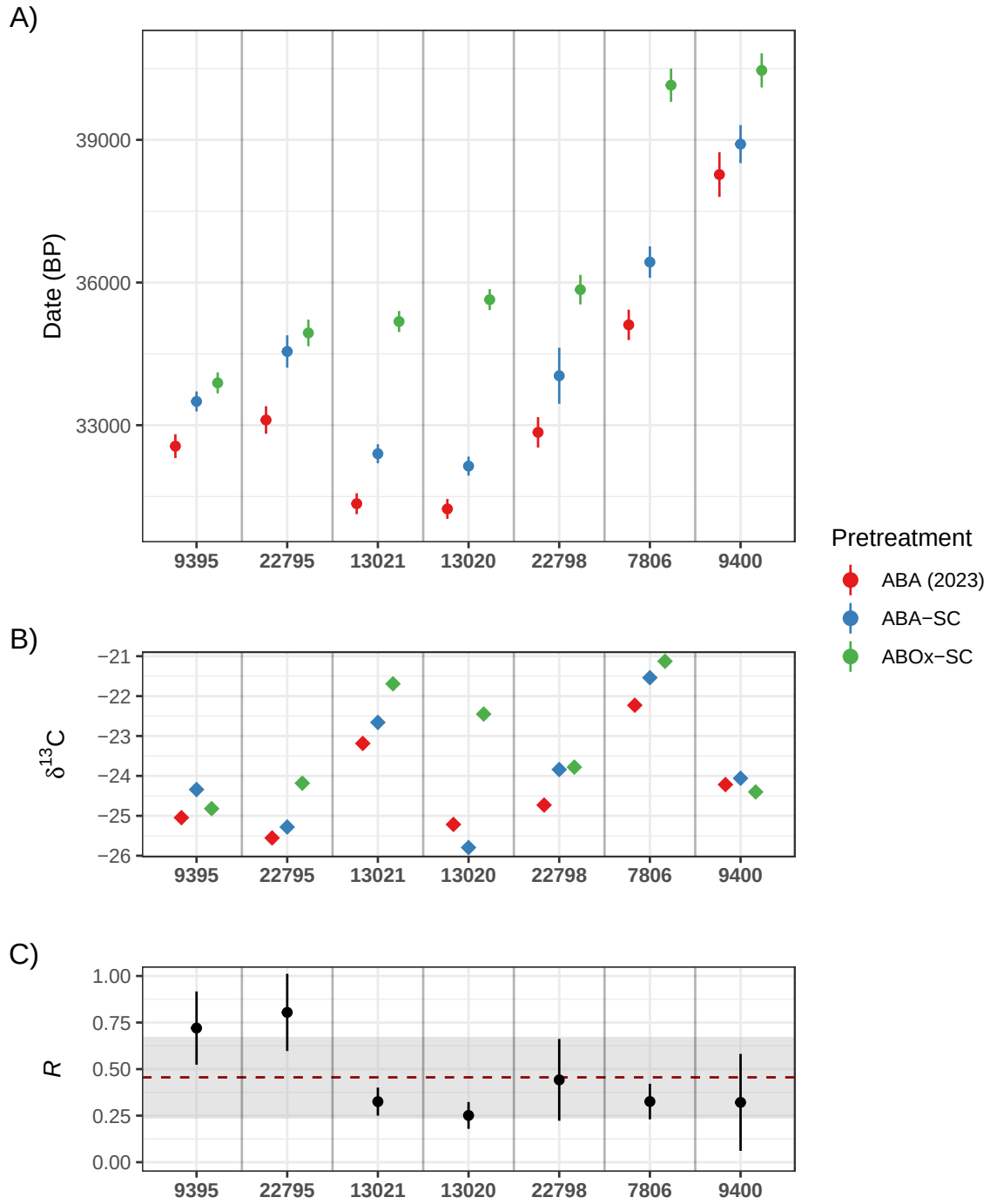


Figure 5.5: Summary of new ABA-SC radiocarbon determinations from Fumane, with ABA and ABOx-SC dates for comparison. Panel A shows radiocarbon ages in ^{14}C years for each sample for ABA, ABA-SC, and ABOx-SC pretreatment methods. Panel B shows stable carbon isotope ratios for the samples and methods. Panel C shows the proportional effect of precombustion on ABA samples for contaminant removal, relative to the inferred amount removed by ABOx-SC (the higher the effect, the more contamination was removed, expressed as a proportion of the amount removed by ABOx-SC). Red dashed line and grey shading in Panel C show the mean and standard deviation across samples for R .

Table 5.4: The effect of precombustion on ABA-treated samples. Column headings: SC'd = amount of material precombusted (mg); SC Yield = amount of material remaining after precombustion (mg); % Change = percentage mass change; $\Delta\%C$ = change in percent carbon measurement after precombustion; $\Delta\delta^{13}C$ = change in stable carbon isotope ratio after precombustion; χ^2 = Chi-squared metric for ABA and ABA-SC dates ($df = 1$, $p < 0.05$); R = proportional effect of precombustion on ABA samples for contaminant removal, relative to the inferred amount removed by ABOx-SC.

Sample	Context	SC'd	SC Yield	% Change	$\Delta\%C$	$\Delta\delta^{13}C$	χ^2	R	σ_R
Fumane									
9395.0	D3ba	15.94	4.12	-74.15	30.38	0.7	8.29	0.75	0.06
22795.0	A2	13.77	2.25	-83.66	19.99	0.3	10.38	0.83	0.06
13021.0	A2	16.39	6.62	-59.61	26.64	0.5	12.47	0.32	0.01
13020.0	A2	22.00	11.37	-48.32	27.41	-0.6	9.63	0.25	0.01
22798.0	A2	8.96	1.49	-83.37	15.87	0.9	3.14	0.46	0.06
7806.0	A5	12.05	2.76	-77.10	32.29	0.7	8.25	0.33	0.01
9400.0	A5/A6	15.94	7.39	-53.64	24.98	0.1	1.08	0.32	0.07
Maninjau									
22401.5		7.44	3.26	-56.18	6.27	0.8	0.06		
22401.6		10.54	6.71	-36.34	-0.39	1.6	0.29		

five out of the seven archaeological samples, the effect of stepped combustion was statistically significant, indicated by χ^2 values for ABA and ABA-SC dates greater than 3.84 ($df = 1$, $N = 2$, $p = < 0.05$, Table 5.4).

Two Fumane samples showed χ^2 scores for ABA and ABA-SC less than the critical value of 3.84: 22798 and 9400. The AMS measurement error for sample 22798 was relatively high for the ABA-SC date (± 590 ^{14}C years). This is due to a low number of AMS measurements made (2 rounds), further problematised by a low mean current (16.4 μA) over rounds. These factors contributed to a higher-than-expected error, which in turn produced a low χ^2 value, despite the measured dates for 22798 for ABA and ABA-SC being 1190 ^{14}C years apart (Figure 5.5, Table 5.4).

%C measurements for the three pretreatment methods considered are also shown in Table 5.3 and summarised in Figure 5.6. These results indicated a higher spread for ABOx-SC ($M = 66.08$, $SD = 13.2$) than for ABA ($M = 66.04$, $SD =$

8.16) and ABA-SC ($M = 86.42$, $SD = 5.26$). ABA-SC showed considerably higher %C values than other methods, including ABOx-SC, likely due to differences in precombustion methods. Only the ABOx-SC method for sample 22798 showed an anomalously low %C yield ($< 50\%$). The lower-than-expected %C value for this sample is not obviously explained by any laboratory observations or notes made during chemical pretreatment, but may have been due to sediment adhering to the sample that was not removed during sampling or wet chemistry.

None of the $\delta^{13}\text{C}$ measurements made fell outside of the range expected values for charcoal, but some samples showed variability in $\delta^{13}\text{C}$ across pretreatment methods (Figure 5.5). Assuming $\delta^{13}\text{C}$ values can be expected to be consistent *within* charcoal samples (i.e., no variability would be observed in $\delta^{13}\text{C}$ between different regions of the same sample), the variable $\delta^{13}\text{C}$ results *across* methods tentatively suggested that each pretreatment method (ABA, ABOx-SC, or ABA-SC) removed a distinct contaminant fraction, with a distinct isotopic signature.

Further data relating to changes in mass, %C, and $\delta^{13}\text{C}$ values, as observed before and after precombustion is given in Table 5.4. Precombusting ABA material resulted in percentage decreases in mass between ~ 85 and 30% . However, starting weights for precombustion were variable, and often were below the recommended sample weight for precombustion (20 mg, Brock et al., 2010).⁶ Changes in %C yields before and after precombustion are all positive, except the Maninjau sample 22401.6. This sample showed a negligible negative change, within the bounds of measurement error. Except 13020 (-0.6), all changes in $\delta^{13}\text{C}$ were also positive at between 0.1 and 1.6, suggesting an overall enrichment of $\delta^{13}\text{C}$ by precombustion.

Although improved removal of contamination can be inferred from ABA-SC results for all Fumane samples, the method was unable to reproduce dates as old as those by ABOx-SC. This suggests that the aggressive pretreatment chemistry of ABOx-SC (i.e., using increased acid and base concentrations, and/or oxidation

⁶In interpreting percentage changes in mass after precombustion, it is important to note that weighing samples at this stage of the pretreatment process can be difficult, particularly for smaller precombustion yields. Static activity in quartz glass combustion tubes often inhibits accuracy during weighing and can result in sample loss, which is more significant for % mass changes in smaller samples.

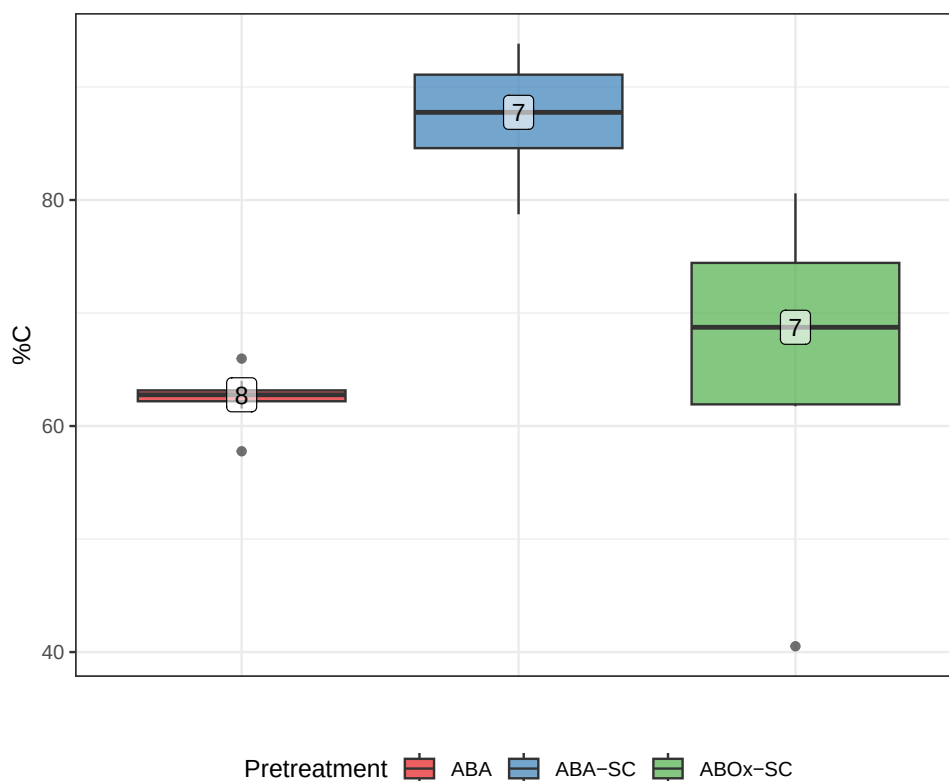


Figure 5.6: Boxplots showing %C measurements for ABA, ABA-SC, and ABOx-SC. Numbers over boxplots indicate the number of samples summarised for each pretreatment.

by acidified potassium dichromate) in combination with precombustion improved the removal of exogenous carbon from charcoal material.

5.3.2 Proportional Effect of Precombustion on Contaminant Removal

The effect of precombustion on contaminant removal for ABA-treated material, expressed as a proportion of the amount removed by ABOx-SC (R , Section 5.2.5), is shown in Figure 5.5C and Table 5.4. On average, ABA-SC improved contaminant removal over ABA by 46.5% ($\sigma = 23\%$).

5.3.3 Duplicate ABA Measurements and the Effect of Long-term Storage at the ORAU

Given that each sample has at least one previous ABA date to the one produced in this study, an opportunity to assess the effect of long-term storage at the ORAU is

Table 5.5: Original and updated (2023) radiocarbon measurements for samples pre-treated by ABA.

Sample	Context	ABA (Original)					ABA (2023)				χ^2
		Year	%C	$\delta^{13}\text{C}$	Date	$\pm$	%C	$\delta^{13}\text{C}$	Date	$\pm$	
Fumane											
9395.0	D3ba	1998	62.24	-24.82	32020	340	62.41	-25.05	32560	250	1.64
22795.0	A2	2008	60.69	-25.55	32480	330	64.01	-25.21	33110	290	2.06
13021.0	A2	2002	44.31	-23.30	31830	260	62.75	-22.96	31350	220	1.99
13020.0	A2	2002	58.47	-25.22	30650	260	57.77	-25.68	31240	210	3.12
22798.0	A2	2008	64.37	-24.74	33320	330	62.87	-24.72	32850	320	1.05
7806.0	A5	1996	62.30	-22.09	33700	600	61.55	-22.23	35110	320	4.30
9400.0	A5/A6	1998	42.06	-23.80	38800	750	62.78	-24.22	38270	470	0.36
Maninjau											
22401.5		2008	81.19	-27.87	48700	900	80.36	-27.77	49400	2300	0.08
22401.6		2008	82.43	-25.18	50500	1000	79.83	-24.93	48700	1400	1.09

inherent in the experimental design. In some cases, it has been over 25 years since an ABA date was first made. By comparing old and new dates, the degree to which samples absorbed modern atmospheric carbon after storage, if any, was measured. Detailed data for these comparisons is shown in Table 5.5, and plotted in Figure 5.7.

Comparisons between old and new dates showed some differences in ^{14}C measurements, but no consistent pattern across samples. Other than 7806, the paired dates were statistically indistinguishable, with χ^2 values less than 3.84 ($\text{df} = 1$, $N = 2$, 5%). Except for Maninjau samples, lower measurement error was achieved with updated results, most likely due to the improved precision of the updated AMS at the ORAU.

Except for samples 9400 and 13020, %C values were consistent across the duplicated ABA measurements, all within 5% difference. The differences observed for 9400 and 13020 were unexplained, but could both be due to a weighing error before combustion. New $\delta^{13}\text{C}$ values were very consistent with previous values, within 0.5 parts per thousand difference for all samples.

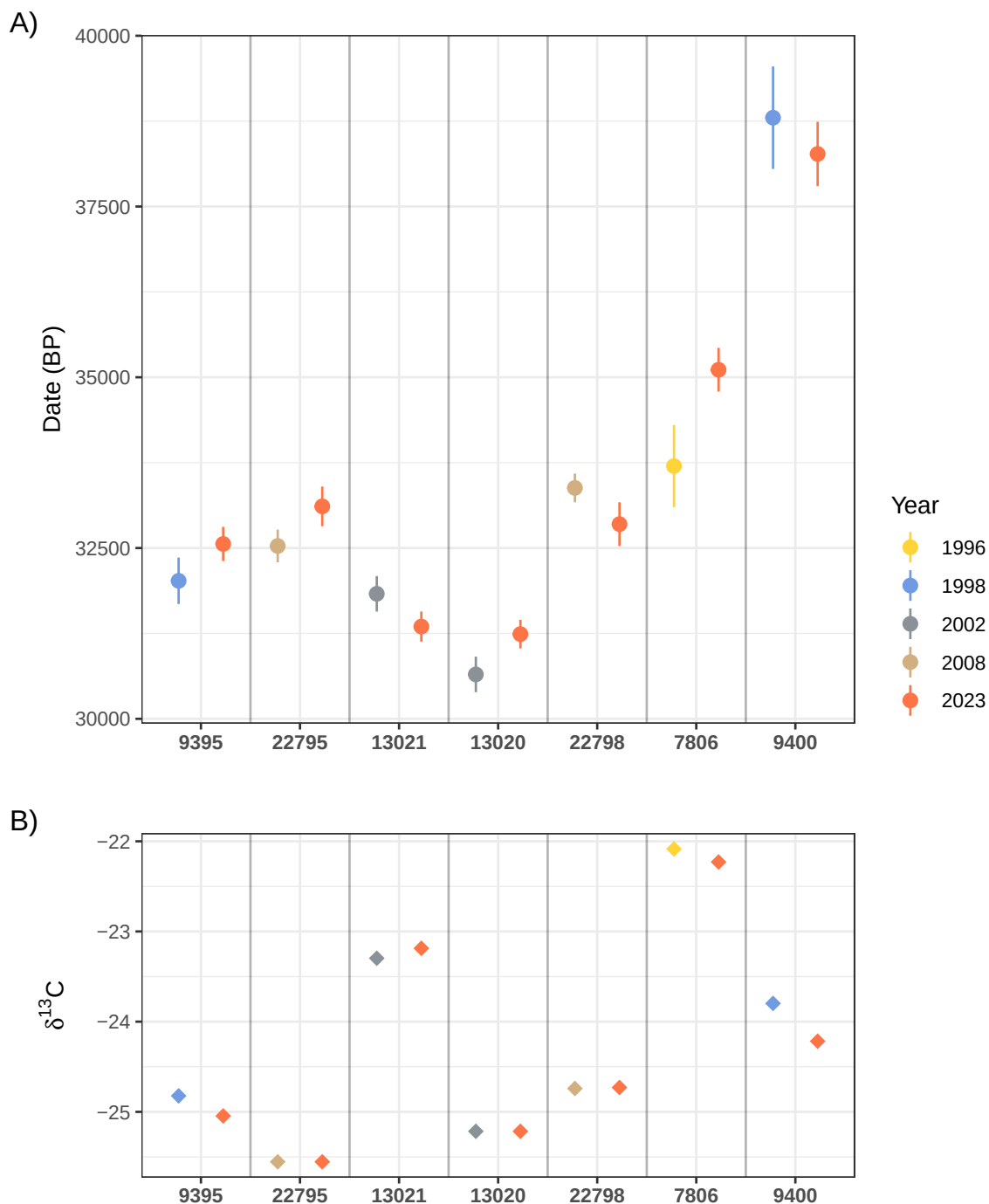


Figure 5.7: Comparison of duplicate ABA radiocarbon and stable isotope measurements for Fumane samples made at the ORAU since 1996. Panel A shows radiocarbon measurements for each sample in radiocarbon years. All measurements made in 2023 (purple) were made for this experiment. Panel B shows stable carbon isotope ratios.

5.4 Discussion

The experiment presented in this chapter allows a detailed consideration of the effect of precombustion in pretreatment of charcoal samples for radiocarbon dating. Specifically, the experiment aimed to quantify the effect of precombustion in removing exogenous carbon from charcoal samples, and investigate whether pretreatment by ABA-SC can produce radiocarbon dates similar to pretreatment by ABOx-SC. This was achieved by a relatively simple experiment design, involving the precombustion and re-dating of excess material previously pretreated by ABA at the ORAU. Archaeological samples used in the experiment were from Grotta di Fumane, a cave site in north-eastern Italy, where charcoal throughout the site's sequence has previously been shown to be contaminated to varying degrees, indicated by significant differences in previously produced ABOx-SC and ABA dates (Brock & Higham, 2009; Higham, Brock, et al., 2009).

The pattern in ABA, ABA-SC, and ABOx-SC dates is very clear from the seven samples dated (see Figure 5.5 and Section 5.4.1). This experiment represents the largest set of radiocarbon dates for charcoal pretreated by ABA, ABA-SC, and ABOx-SC to date. Previous studies directly comparing ABA-SC and ABOx-SC were either limited by low numbers of paired samples (e.g. Bird et al., 1999; Tomiyama et al., 2016), or by material type (Santos et al., 2001; Santos & Ormsby, 2013; Turney, Bird, et al., 2001), and most comparisons between ABA-SC and ABOx-SC were inconclusive (Table 5.1). The experiment presented here shows irrefutably that the less aggressive ABA protocol combined with precombustion is insufficient to remove contamination, and likewise that the aggressive chemical protocols of ABOx-SC are - at least in the case of charcoal from Fumane - essential for removing it (Table 5.3).

5.4.1 Radiocarbon Measurements and Contamination Removal by Precombustion

This study confirms that precombustion can contribute significantly to contaminant removal. Having similarly observed that stepped combustion systematically

produced older dates for both ABA and ABOx, [Tomiyaama et al. \(2016\)](#) discuss a number of possible sources for carbon contaminants that stepped combustion specifically removes: contaminant carbon introduced by copper oxide used in combustion rigs ([Vandeputte et al., 1998](#)); atmospheric CO₂ or modern carbon introduced during pretreatment ([Bird et al., 1999](#)); or “organic remnants after chemical treatment” ([Tomiyaama et al., 2016: 577](#)). Results from this experiment suggest precombustion removes the latter, organic contaminants present in the sample. Given that copper oxide was not used in precombustion or combustion in this experiment ([R. Wood & Brock, 2011](#)), and no atmospheric CO₂ absorption or laboratory contamination was evident from Mananjau standards (Table 5.3), contaminants removed by precombustion after wet chemistry must be organic remnants of environmental carbon, with origin in the sediment column at Fumane.

Although no samples in this experiment were pretreated by ABOx wet chemistry only (i.e., ABOx without precombustion), previous studies have highlighted the significant effect of stepped combustion in ABOx-SC ([Bird et al., 2014](#); [Tomiyaama et al., 2016](#)). Results from this chapter’s experiments, as well those in the following chapter, confirm that dates are typically older after precombustion compared to before, regardless of the wet chemistry protocol used. This adds further evidence to the conclusion that precombustion is an essential step in the ABOx-SC protocol.

The interaction between the precombustion process and the various chemical stages of both the ABA and ABOx chemical protocols are more complex, but equally important questions. The combined effect of aggressive oxidation and stepped combustion is clearly substantial, given that ABA-SC was unable to reproduce ABOx-SC dates, and ABOx without stepped combustion was unable to produce dates as old as ABOx-SC in other published experiments ([Bird et al., 2014](#); [Tomiyaama et al., 2016](#)). [Tomiyaama et al. \(2016\)](#) argue that base washes during ABA or ABOx protocols can remove low-stability exogenous components from a sample, but that residues may remain after wet chemistry, tightly bound to the edges of aromatic structures of the charcoal. The improved contaminant removal by ABA-SC may be due to an interaction between base washes and

precombustion: the base wash may partially degrade contaminant bonds, increasing contaminant lability, and these are broken during precombustion. Oxidation under ABOx-SC may further degrade these bonds, but a downside of this is aggressive protocol is that these processes can also degrade aromatic compounds endogenous to the charcoal sample itself, as shown by [Ascough, Bird, Francis, & Lebl \(2011\)](#) (see Chapter 6).

Some dates produced for this experiment show relatively high errors and low mean AMS current (e.g., 22798, $< \sim 20 \mu\text{A}$, Table 5.3). This is probably due to a number of mistakes made during combustion and graphitisation that meant graphite yield was low, and the number of AMS rounds made was less than usual. Unfortunately, there was not enough excess material to re-date these samples. Although these samples' large errors have some effect on respective χ^2 values, their measurements do not deviate from the overall pattern of results (Figure 5.5).

5.4.2 %C and $\delta^{13}\text{C}$ Measurements for ABA-SC

Comparisons between %C measurements for ABA, ABA-SC, and ABOx-SC show a clear pattern: ABOx-SC generally had higher or equivalent %C yields than ABA, but ABA-SC was consistently highest, with much less spread in ABA and ABA-SC values than for ABOx-SC (Figure 5.6, Table 5.4). Positive changes in %C values were observed for all ABA samples from Fumane after precombustion, with increases of between ~ 15 and 35% (Table 5.5).

Typical %C values for charcoal are between 60% and 80%, and many argue that %C values less than 50% should be cautiously interpreted in radiocarbon dating, if not rejected ([Rebollo et al., 2011](#); [Tomiyaama et al., 2016](#); [R. Wood, 2015](#)). Low %C values for charcoal may indicate that substantial contamination is present in a sample, that may also be a source of exogenous carbon. At the ORAU, %C values are the only independent measurement that might indicate this. Only one measured %C value from Fumane falls below 50%, the ABOx-SC treated sample 22798, at 40.51 %C. No obvious reason for this was found in laboratory notes or relevant publications ([Brock & Higham, 2009](#); [Higham, Brock, et al., 2009](#)), and

much higher %C values for ABA and ABA-SC were observed for the same sample (62.87% and 78.74%, respectively).

Lower %C values for the ABOx-SC sample 22798 may be due to the aggressive nature of the method's wet chemistry that promotes self-humification in charcoal (Alex et al., 2017; Boaretto et al., 2008; Cohen-Ofri et al., 2006). Depending on the stability and preservation of the sample, the high-concentration base washes and aggressive oxidation by acidified potassium dichromate may have removed substantial amounts of charcoal from the bulk sample, but fail to remove non-carbonaceous components, such as sediment. Regardless of how much sediment was included in the original sample, if enough charcoal was removed during wet chemistry, a reduction in %C would likely be observed, as the %C content of sediment is typically much lower than charcoal. This might also explain the wider spread of ABOx-SC %C values, which for ABA and ABA-SC are comparatively consistent (Figure 5.6).

The spread %C values for ABOx-SC samples might simply be due to excessive sediment adherence that was overlooked for some ABOx-SC samples, but not for ABA. This might explain anomalously low results such as 22798, but fails to account for consistently higher ABA-SC %C values (Figure 5.6). A more likely suggestion is that ABOx-SC %C values included materials introduced during precombustion. The precombustion method at the ORAU was changed in 2012 (see Section 5.2.3), *after* all ABOx-SC dates compared in this study were made (R. Wood & Brock, 2011). Prior to this, ~ 200 mg copper oxide and a plug of quartz wool were used in precombustion (Brock & Higham, 2009).

The change in protocol at the ORAU followed an internal report, produced in light of possible evidence of contamination introduced by copper oxide suggested by Vandeputte et al. (1998). No evidence for contamination by copper oxide was found, but the report showed through FTIR and radiocarbon measurements that precombustion using oxygen produced indistinguishable results to the copper oxide method, and as a precaution the precombustion protocol was changed. It is possible that copper oxide or quartz wool was consistently included in ABOx-SC

sub-samples when weighed for combustion, affecting %C yields. This is not likely to have had an impact on ^{14}C measurements, as both materials were heated at 850 °C for 8 hr before precombustion and would be “inert” in terms of their carbon content (Brock & Higham, 2009).

Except for sample 13020, changes in $\delta^{13}\text{C}$ values show consistent enrichment following precombustion (Table 5.4), likely due to fractionation during precombustion where carbon and oxygen react to form CO_2 and the lighter $\delta^{13}\text{C}$ isotope is lost preferentially (Table 5.4). Another interpretation is that the carbon fraction made more labile by pretreatment, a proportion of which is removed during precombustion (i.e., the contaminant fraction), has a distinct and lighter isotopic signature compared to the endogenous fraction. One way to examine this in more detail would be to capture CO_2 released during lower temperature precombustion, and compare its isotope ratio to the high temperature fraction, similar to the original stepped combustion study by Bird et al. (1999), and later studies by Ascough et al. (2008). This in combination with *known* contaminant experiments (e.g., Bird et al., 2014, 2017; Orr et al., 2021; Tilston et al., 2016) would provide more insight on this topic (see Chapter 7).

The broader differences between $\delta^{13}\text{C}$ measurements for ABOx and ABA, with or without precombustion, are more complex questions. Clusters of $\delta^{13}\text{C}$ values for each sample show significant variation, with no clear pattern in results (Figure 5.5). All but one ABA sample showed enrichment after precombustion, with four out of the seven samples showing further enrichment for ABOx-SC. In all but one of ten samples, R. Wood et al. (2012) also reported enriched $\delta^{13}\text{C}$ values for ABOx-SC treated samples compared to ABA. Again, it is possible that this is simply due to fractionation during the precombustion step of ABOx-SC. However, given that the majority of results here showed further enrichment for ABOx-SC compared to ABA-SC, it is possible that some fractionation also occurs during the oxidation step of ABOx-SC.

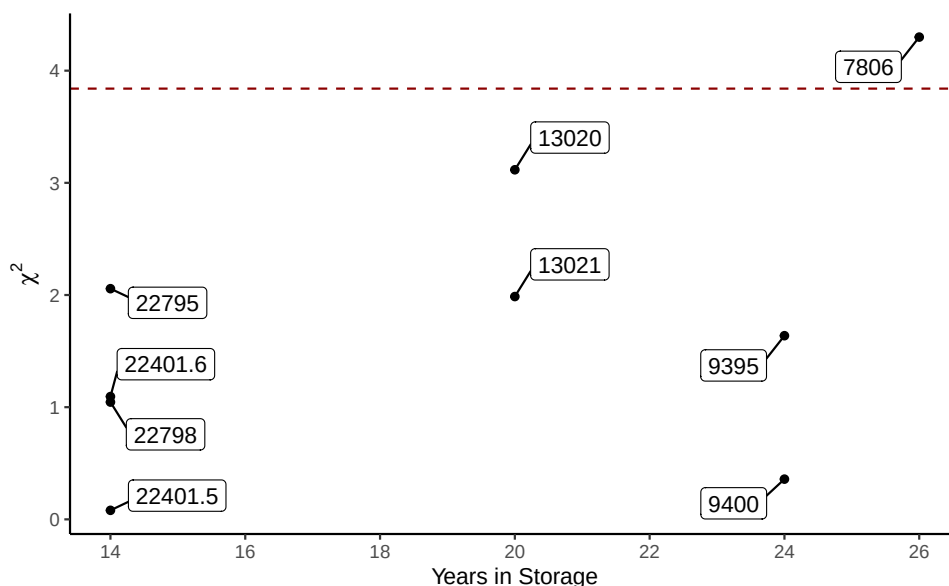


Figure 5.8: Paired χ^2 values for original and updated ABA radiocarbon dates (y-axis) against the number of years between measurements (Years in Storage, x-axis). Dashed horizontal line shows the value above which χ^2 values are significant (3.84, $df = 1$, $N = 2$).

5.4.3 Long-term Storage of Excess Material at the ORAU

The effect of long-term storage on excess pretreated charcoal at the ORAU was also assessed as a result of the updated (i.e., duplicated) ABA measurements made as part of this study. A concern here was that the new dates would be systematically younger than the original dates, indicating that samples had absorbed modern contaminant carbon during their storage. Figure 5.8 and Table 5.5 show that in all but one case (7806), no significant differences between ^{14}C measurements were observed, while the new measurement for 7806 is in fact significantly older than the original performed in 1996.

$\delta^{13}\text{C}$ values for old and new measurements show remarkable consistency, with the maximum difference between old and new measurements being within the range of instrument error. This is reassuring for the reliability of measurements used in this study, as well as for consistency in stable isotope measurements over the last 25 years at the ORAU.

%C values are also very consistent, with the exception of samples 9400 and 13021, whose original %C values were much lower than the most recent ones.

These anomalous measurements, made in 1998 and 2002, respectively, are difficult to explain. Again, no archival laboratory notes indicate any obvious reasons for lower %C yields, and neither were %C values unusual for the same samples after precombustion for ABA-SC (Table 5.3). The original weights for these samples were fairly high compared to the 3 mg recommended weight currently used at the ORAU (4.7 mg for 9400 and 4.8 mg for 13021), perhaps indicating incomplete freeze drying. Other explanations are weighing or recording errors before combustion. However, for both samples, any error would have had to have been fairly large to have had such an effect, approximately 1.5 mg for 9400 and 1.4 mg for 13021, assuming their updated %C values are correct.

If their original %C measurements for 9400 and 13021 in fact accurate, important questions are raised concerning the consistency of %C that can be expected within samples. It may be the case that Bird & Ascough (2012)’s “combustion continuum” also applies *within* samples (i.e., stability and susceptibility to contamination vary across different regions of a single charcoal sample), as well as between them, for example.

Excess charcoal remaining after pretreatment at the ORAU is typically stored directly in microtubes, or folded in aluminium foil parcels that are then stored in a microtube. Revisiting samples stored in aluminium foil revealed that in some cases the foil was partially dissolved. This was probably due to residual HCl on the charcoal that reacted with the aluminium and partially degraded it. The partially dissolved foil was very fragile, and was difficult to remove when re-sampling for updated ABA and ABA-SC measurements. It is recommended that samples treated with HCl are stored directly in microtubes in future, without further wrapping in aluminium foil. The downside of this is that if excess yields are very small, static in the microtube can be very difficult to control, and very fine powdered or small samples can easily be lost during weighing or sample handling.

5.5 Conclusion

The experiment presented in this chapter was designed to investigate the effect of precombustion as a pretreatment step in radiocarbon dating contaminated charcoal samples. The results show that, in the majority of cases, precombusting ABA material produces significantly older radiocarbon dates than ABA, but that ABA-SC is unable to remove contamination as effectively as ABOx-SC. By extension, this suggests that precombustion is an essential stage of the ABOx-SC pretreatment method. The results from this experiment raise further questions around the role of strong acid and base washes in ABOx-SC, as well as the aggressive oxidation by acidified potassium dichromate, and their effect in producing significantly older dates than any other pretreatment method.

In the following chapter, an alternative oxidation reagent is tested with additional samples from Fumane, providing further insight into the effect of the aggressive chemical protocols of ABOx-SC. The combined findings from the two experiments and directions for future research are then discussed in Chapter 7.

6

Acid-Bleach with Stepped Combustion (ABl-SC) as a Pretreatment Method for Radiocarbon Dating of Charcoal

Abstract

In many contexts, chemical pretreatment methods have been shown to have considerable effect on radiocarbon dates. Chapter 5 explored an alternative pretreatment method for charcoal, demonstrating that adding a precombustion step to the ABA protocol (ABA-SC) significantly improves removal of contamination before the radiocarbon dating of charcoal. However, results indicated that precombustion without a preceding oxidation step in chemical pretreatment is insufficient to remove contaminants to the same level as ABOx-SC.

The ABOx-SC protocol uses potassium dichromate as an oxidation reagent, a highly toxic substance that requires careful handling in a laboratory setting. Regularly processing ABOx-SC samples using potassium dichromate at the ORAU is time consuming and requires extensive measures to avoid harm to human and environmental health. This chapter presents an experimental pretreatment method with sodium chlorite (NaClO_2) as an alternative and less toxic oxidation reagent. As for the previous chapter, samples from Middle and Upper Palaeolithic layers

at Grotta di Fumane were used to test the experimental method, again that have previously been shown to produce significant differences between dates produced by ABA and ABOx-SC. Three of these samples were also used to test the effect of precombustion by ABA-SC in Chapter 5. In this study, samples were also dated with and without precombustion, allowing further consideration of the effects of precombustion.

Results indicate that oxidation by sodium chlorite (ABl) is unable to reproduce dates as old as ABOx-SC, but generally show improvement over ABA. As for the previous chapter, precombustion improved contaminant removal further for ABl-pretreated samples. For the three samples dated in both experiments, ABl-SC also showed improvement over ABA-SC, with only one ABl-SC date comparable to ABOx-SC. Overall, results suggest that it is the combined effect of aggressive oxidation and precombustion under ABOx-SC that contributes to more effective contaminant removal. It is also possible that the high-strength base step in ABOx-SC used to date Fumane samples in 2007 and 2008 had an effect on contaminant removal, but because the base step was omitted from the experimental protocol here, this is difficult to quantify without further investigation.

6.1 Introduction

As demonstrated and discussed in Chapter 5, pretreatment chemistry and laboratory protocols can have considerable effect on radiocarbon dates, particularly when dating material produced close to the radiocarbon boundary (50 ky BP, [R. Wood, 2015](#)). The experiment presented in the previous chapter aimed to quantify the effect of precombustion as a pretreatment step in the ABOx-SC method for radiocarbon dating of charcoal, using an adapted ABA protocol with precombustion (ABA-SC). By comparing dates produced by ABA, ABA-SC, and ABOx-SC, the proportional effect of precombustion after ABA was quantified at around 40% of the relative contamination removed by ABOx-SC, but in some cases was as high as 80%. In other cases, only 25-35% could be attributed to precombustion. While precombustion is clearly an important aspect of the ABOx-SC protocol, it is only part of the story. The effect of other components of ABOx-SC remain unquantified; it is still unclear why the ABOx-SC method can so effectively remove contaminants in comparison to ABA.

Details of the sources and process of exogenous contamination in charcoal, an overview of methods for their removal such as ABA and ABOx-SC, and descriptions of the protocols and precombustion method are given in Chapter 5.

As previously stressed, several key differences in charcoal pretreatment were introduced by [Bird et al. \(1999\)](#)'s ABOx-SC method, in comparison to ABA. To reiterate, these were:

1. Increasing the concentrations of the acid and base solutions
2. Replacing the second acid wash with an oxidation step, using acidified potassium dichromate
3. Using stepped combustion before graphitisation and AMS measurement

Considering these three differences in light of the results from experiments in ABA-SC in Chapter 5, the aggressive oxidation step, increased acid and base wash concentrations, or their effect in combination with precombustion, are likely

to comprise the remaining contribution to improved contaminant removal under ABOx-SC. This chapter focuses on the second point of the above, by exploring the effect of the oxidation step in ABOx-SC.

Oxidation in ABOx involves the use of potassium dichromate ($K_2Cr_2O_7$), a highly toxic and carcinogenic substance that requires careful handling and extensive measures during its laboratory application to avoid harm to human and environmental health. The practicalities of processing samples for ABOx in a radiocarbon laboratory involve extensive fume hood preparation, extremely careful handling while applying and removing potassium dichromate solutions, and equally careful washing and handling of any waste or equipment that comes in to contact with potassium dichromate (e.g., water used for rinsing samples, paper, aluminium foil, or glassware). Although the ABOx chemical pretreatment has been shown to decontaminate samples more effectively than other methods, as further highlighted by results from Chapter 5, regularly processing samples by ABOx-SC is difficult and at times impractical within the laboratory setting, simply due to the measures required when handling potassium dichromate. An investigation into alternative oxidation methods that are less harmful to humans and the environment, and require less laborious preparation in their application, is worthwhile and necessary.

Because the ABOx-SC method involves relatively aggressive chemical washes, its rate of failure is considerably higher than for ABA.¹ Samples are frequently lost during the base wash, where the extreme change in pH from 6M HCl to 1M NaOH (pH -0.77 to pH 14) can destroy the majority of the charcoal sample and leave insufficient material for subsequent pretreatment stages. As discussed in Chapter 5, charcoal samples can undergo significant oxidation and “self-humification” after deposition to the burial environment, meaning that they are particularly susceptible to failure during the base step (Ascough, Bird, Francis, Thornton, et al., 2011; Cohen-Ofri et al., 2006, 2007). Although a greater amount of starting material

¹“Failure” in radiocarbon dating can be due to many reasons and can describe many scenarios, but here refers to samples that do not survive (i.e., are completely dissolved during) pretreatment chemistry. Reasons for failure might relate to sample size, fragility, depositional context, or degree of contamination.

is recommended in order to obtain sufficient material for graphitisation, samples are often destroyed completely by the protocol's aggressive chemistry (Bird et al., 2010; Sutikna et al., 2016; R. Wood, 2015).

For this reason, the ORAU began to implement an adjusted ABOx-SC method in 2016, omitting the base step from the protocol. Internal experiments showed that Acid-Oxidation with stepped combustion (AOx-SC) produces dates indistinguishable to the ABOx-SC method. Currently these experiments are unpublished, but the AOx-SC method is now routinely applied at the ORAU, and a number of Palaeolithic radiocarbon chronologies have been published using the method (Douka et al., 2019; Loftus et al., 2019; Pargeter et al., 2018; Sutikna et al., 2016). The AOx-SC protocol as implemented at the ORAU is identical to ABOx-SC in terms of concentration and duration of acid and oxidation washes, as well as precombustion, the only difference is the omission of the base step.

Sodium chlorite (NaClO_2) is an oxidation reagent routinely used at the ORAU at varying concentrations, applied during protocols for the chemical pretreatment of skin, parchment, leather, plant remains, and wood (Brock et al., 2010). Sodium chlorite does not require any special laboratory preparation for use, other than handling under a fume hood. If sodium chlorite can be used as an oxidation reagent in AOx-SC or ABOx-SC instead of potassium dichromate and produce similar results, laboratory protocols at the ORAU could be greatly simplified.

As well as the practical implications of potentially simplifying laboratory methods, an experiment using an alternative and less aggressive oxidation reagent - with and without precombustion - could further illuminate the mechanisms by which ABOx-SC and AOx-SC often produce older dates in certain contexts. As suggested above, it may be that the strong and aggressive chemical oxidation step in ABOx-SC *in combination with* precombustion has a significant effect on dates. Experimental data using a less aggressive oxidation reagent could confirm or further illuminate this question.

This chapter presents a second experiment in the radiocarbon dating of charcoal, replacing acidified potassium dichromate in AOx-SC with acidified sodium chlorite

as an oxidation reagent (ABL-SC). As for Chapter 5, results with and without precombustion are compared. In doing so, the following questions are considered:

1. *Can Acid-Sodium chlorite oxidation with or without precombustion (ABL-SC) produce radiocarbon dates similar to ABOx-SC?*
2. *Does sodium chlorite oxidation in combination with precombustion affect %C or $\delta^{13}\text{C}$ measurements, in comparison to ABOx-SC?*
3. *Does ABL-SC introduce any background contamination during the pretreatment process?*

6.2 Materials and Methods

Table 6.1: Archaeological samples used in experiment from Grotta di Fumane, previously dated by ABA and ABOx-SC methods. χ^2 statistics for paired dates are given in the final column ($df = 1$, $N = 2$, $p < 0.05$).

Sample	Industry	Context	ABA				ABOx-SC				χ^2
			$\delta^{13}\text{C}$	%C	Result	$\pm$	$\delta^{13}\text{C}$	%C	Result	$\pm$	
13022	Gravettian	D1d	-23.4	55.4	31490	250	-22.3	78.6	31590	160	0.11
22799	Proto Aurignacian	A2	-24.5	44.2	32060	360	-24.7	48.6	34100	400	14.37
22798	Proto Aurignacian	A2	-24.7	62.9	32850	320	-23.8	40.5	35750	600	18.19
7806	Mousterian	A5	-21.2	60.7	36860	270	-21.1	74.4	40050	550	27.11
9400	Mousterian	A5 / A6	-24.4	66.0	38710	720	-24.4	62.1	40460	360	4.73

6.2.1 Samples

As for the previous chapter, archaeological samples used in this experiment were from Grotta di Fumane (Fumane). Samples are listed in Table 6.1. A brief description of the samples' stratigraphic context and the wider archaeological significance of Palaeolithic layers at Fumane is provided in Chapter 5, and discussed in more detail by Broglio et al. (2004), Peresani et al. (2008), and Higham, Brock, et al. (2009). Samples were from Middle Palaeolithic (Mousterian; layers A5 and A6) and Upper Palaeolithic (Proto Aurignacian and Gravettian; layers A2 and D1d) layers. All but one of these samples (13022) have previously produced significant differences for dates by ABA and ABOx-SC pretreatments, indicated by their chi-squared statistic in Table 6.1.

Samples were selected that had sufficient excess untreated material stored at the ORAU for dating by ABl-SC, and that were evenly distributed as possible over the period most relevant to AOx-SC (~ 25 - 50 ky BP). Minimum sample weights followed ABOx-SC and AOx-SC recommendations at the ORAU ($> \sim 100$ mg). Unfortunately, the minimum starting weight requirements for this experiment (> 100 mg) and the requirements of Chapter 5 (having sufficient excess material

Table 6.2: Chemical protocols for ABA, ABOx, AOx, and ABI pretreatment methods. Between all steps samples are washed at least three times with ultra pure water.

	ABA	ABOx	AOx	ABI
1	1M HCl, 80 °C, 20 min	6M HCl, room temperature, 1 hr	6M HCl, room temperature, 1 hr	6M HCl, room temperature, 1 hr
2	0.2M NaOH, 80 °C, 1 hr	1M NaOH, room temperature, 30 min	0.1M K ₂ Cr ₂ O ₇ in 2M HCl, 60 °C, 14 hr	1.5% NaClO ₂ , 60 °C, 24 hr
3	1M HCl, 80 °C, 20 min	0.1M K ₂ Cr ₂ O ₇ in 2M HCl, 60 °C, 14 hr		

after ABA dates for ABA-SC) only allowed for three samples to be dated in both experiments (22798, 7806, and 9400).

In addition to archaeological samples from Fumane, two ¹⁴C-dead standards were also included in the experiment. These were samples of fossilised kauri (*Agathis australis*) wood, from Renton Road, Manukau Harbour, New Zealand (Lorrey et al., 2018). The Renton Road fossils are thought to date at least to Marine Isotope Stage 5, well beyond the radiocarbon limit. The wood was charred at ANU and is routinely used as a background charcoal standard at the ORAU. ¹⁴C determinations on untreated, ABA, AOx-SC treated Renton Road charcoals have all consistently produced ¹⁴C ages greater than 50 ky BP, suggesting the charcoal is exceptionally clean and contaminant-free, making it ideal for use as a laboratory standard. The inclusion of two Renton Road charcoals in the experiment allowed a robust quantification of how much modern contamination is introduced by ABI-SC pretreatment chemistry, if any.

6.2.2 Pretreatment Chemistry and the ABI-SC Method

Fragments of charcoal fragments from bulk samples were hand selected and cleaned of sediment and visible organic matter. The charcoal fragments were roughly crushed before treatment with 6M HCl at room temperature for 1 hr, then rinsing three times with ultra pure water.

For oxidation, a solution of acidified sodium chlorite (1.5% wt/vol) was mixed by adding 7.5g NaClO₂ to 0.06M HCl. The solution was applied to samples for 24 hr at 60 °C. After another three water washes, samples were frozen for 12 hours and freeze-dried for a minimum of 12 hours, as standard at the ORAU (Brock et al., 2010). From the freeze-dried product, ~20 mg was precombusted for ABl-SC, following the precombustion procedure described in Chapter 5. Around 3 mg of the freeze-dried product was weighed for combustion and graphitisation without precombustion (ABl). Samples were weighed after freeze-drying and after precombustion, allowing some inference and comparison of weight loss by different oxidation methods.

Other than replacing potassium dichromate with sodium chlorite as the oxidation reagent and the length of time it is applied to samples, this procedure was exactly the same as AOx pretreatment chemistry (Table 6.2, Brock et al., 2010). Subsequent combustion, stable isotope measurement, graphitisation, and AMS measurement methods were performed as routine at the ORAU.

A schematic diagram of the ABA, ABOx-SC, ABl, and ABl-SC methods is shown in Figure 6.1.

Equations for summarising the improved contaminant removal by ABl and ABl-SC, in comparison to ABA, and as a proportion of contaminant removal by ABOx-SC, are described in Chapter 5. In this chapter, b represents the radiocarbon measurement for either ABl or ABl-SC, and σ_b its associated error.

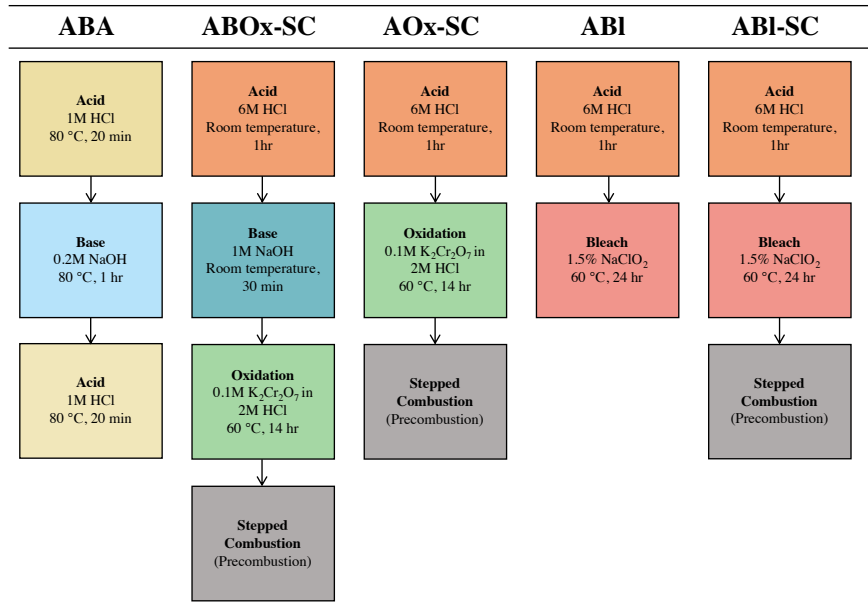


Figure 6.1: Schematic diagram of the ABA, ABOx-SC, AOx-SC, ABl, and ABl-SC pretreatment methods. ABl and ABl-SC are experimental pretreatment methods, whilst ABA, ABOx-SC, and AOx-SC are routinely practiced at the ORAU (Brock et al., 2010).

6.3 Results

Radiocarbon dates and $\delta^{13}\text{C}$ measurements for the five archaeological samples from Fumane are listed in Table 6.3 and plotted in Figure 6.2. Corresponding results for Renton Road, as well as percent carbon values (%C) for all samples are shown in Table 6.3.

Both Renton Road samples' ^{14}C ages were greater than 55 ky BP, indicating that neither ABl or ABl-SC pretreatment methods introduced significant modern contamination to samples before AMS measurement. $\delta^{13}\text{C}$ values for the Renton Road samples were consistent with previous measurements by any pretreatment method at the ORAU, within two standard deviations of the mean ($M = -25.39$, $SD = 0.41$). This suggests that no significant isotopic fractionation occurred during pretreatment chemistry or precombustion. If fractionation occurred during precombustion, significantly enriched $\delta^{13}\text{C}$ values for precombusted samples might be expected, but no such pattern was detectable in the Renton Road standards. Differences in %C measurements for Renton Road samples were also consistent with previous measurements made at the ORAU, with higher % carbon yields for

Table 6.3: Radiocarbon determinations for samples used in experiment. ABl and ABl-SC dates were produced for this experiment, while ABA and ABOx-SC dates were previously produced at the ORAU.

Sample	Treatment	Year	Result	$\pm$	$\delta^{13}\text{C}$	%C
Fumane						
13022	ABA	2002	31490	250	-23.4	55.4
	ABl	2023	31260	270	-24.0	58.0
	ABl-SC	2023	30910	210	-24.1	89.5
	ABOx-SC	2007	31590	160	-22.3	78.6
22799	ABA	2008	32120	240	-24.5	44.2
	ABl	2023	32860	270	-25.6	57.2
	ABl-SC	2023	33140	210	-25.8	83.1
	ABOx-SC	2008	34180	270	-24.7	48.6
22798	ABA	2008	33380	210	-24.7	64.4
	ABA	2023	32850	320	-24.7	62.9
	ABl	2023	33780	410	-24.3	52.6
	ABl-SC	2023	34160	440	-24.1	62.0
	ABOx-SC	2008	35850	310	-23.8	40.5
7806	ABA	1996	33700	600	-22.1	62.3
	ABA	2007	36860	270	-21.2	60.7
	ABA	2023	35110	320	-22.2	61.6
	ABl	2023	36620	450	-21.9	56.8
	ABl-SC	2023	36950	320	-21.8	87.6
	ABOx-SC	2007	40150	350	-21.1	74.4
9400	ABA	1998	38800	750	-23.8	42.1
	ABA	1998	38250	700	-24.2	66.3
	ABA	2007	39500	330	-24.2	60.4
	ABA	2007	39490	350	-24.5	57.8
	ABA	2023	38270	470	-24.2	62.8
	ABA	2023	38710	720	-24.4	66.0
	ABl	2023	39540	620	-25.0	58.5
	ABl-SC	2023	40180	510	-25.0	84.1
	ABOx-SC	2007	40460	360	-24.4	62.1
Renton Road						
49582.16	ABl	2023	56200		-25.1	59.4
	ABl-SC	2023	55600		-24.8	81.1
49582.17	ABl	2023	56300		-24.7	56.2
	ABl-SC	2023	60500		-24.8	78.7

precombusted samples than for those without (ABl-SC vs. ABl; Table 6.3; Figure 6.3). This is typical for precombustion in the presence of O₂ (R. Wood & Brock, 2011), as discussed in Chapter 5.

For all archaeological samples from Fumane, pretreatment by ABOx-SC consistently produced the oldest date. Except for the Gravettian sample 13022², ABl-SC produced the second-earliest dates for all samples. This was followed by ABl, then ABA, for three out of the five samples. Archaeological samples dated by the four different methods showed this pattern for three of the five samples tested (Figure 6.2). Sample 7806 showed the greatest difference between ABOx-SC and all other dates. In this case, ABA, ABl, and ABl-SC were all statistically indistinguishable within 1 σ . Except for sample 13022, all samples pretreated by ABl with precombustion (ABl-SC) produced older dates than without precombustion (ABl), though all paired ABl and ABl-SC dates overlap at 1 σ .

Modern charcoal typically contains between 50% and 70% carbon (Braadbaart et al., 2009). %C yields measured during combustion can be a useful metric through which to infer contamination in charcoal. %C yields less than 50% are a strong indication that material other than charcoal is present in samples, such as sediment. All ABl and ABl-SC samples from Fumane yielded over 50% carbon, suggesting that none of the samples were heavily contaminated with sediment or exogenous material after chemical pretreatment.

As noted in Chapter 5, sample 22798's ABOx-SC treatment yielded 40.5% carbon, indicating that the sample may have included contaminants at the point of combustion. ABl and ABl-SC %C yields for 22798 were both over 50%, but lower than all other samples, suggesting the bulk sample was heavily contaminated, probably by adhering sediment. Given that contamination is much more likely to make radiocarbon dates appear younger than older (R. Wood, 2015), 22798's ABOx-SC date should tentatively be considered a minimum age, which already is considerably older than dates produced by any other method (Figure 6.2).

²It is possible the ABl-SC date for sample 20065 was affected by a reduced amount of iron present in the rig during graphitisation, after a mistake made during rig preparation. This is indicated by a lower than average current measured during AMS measurement.

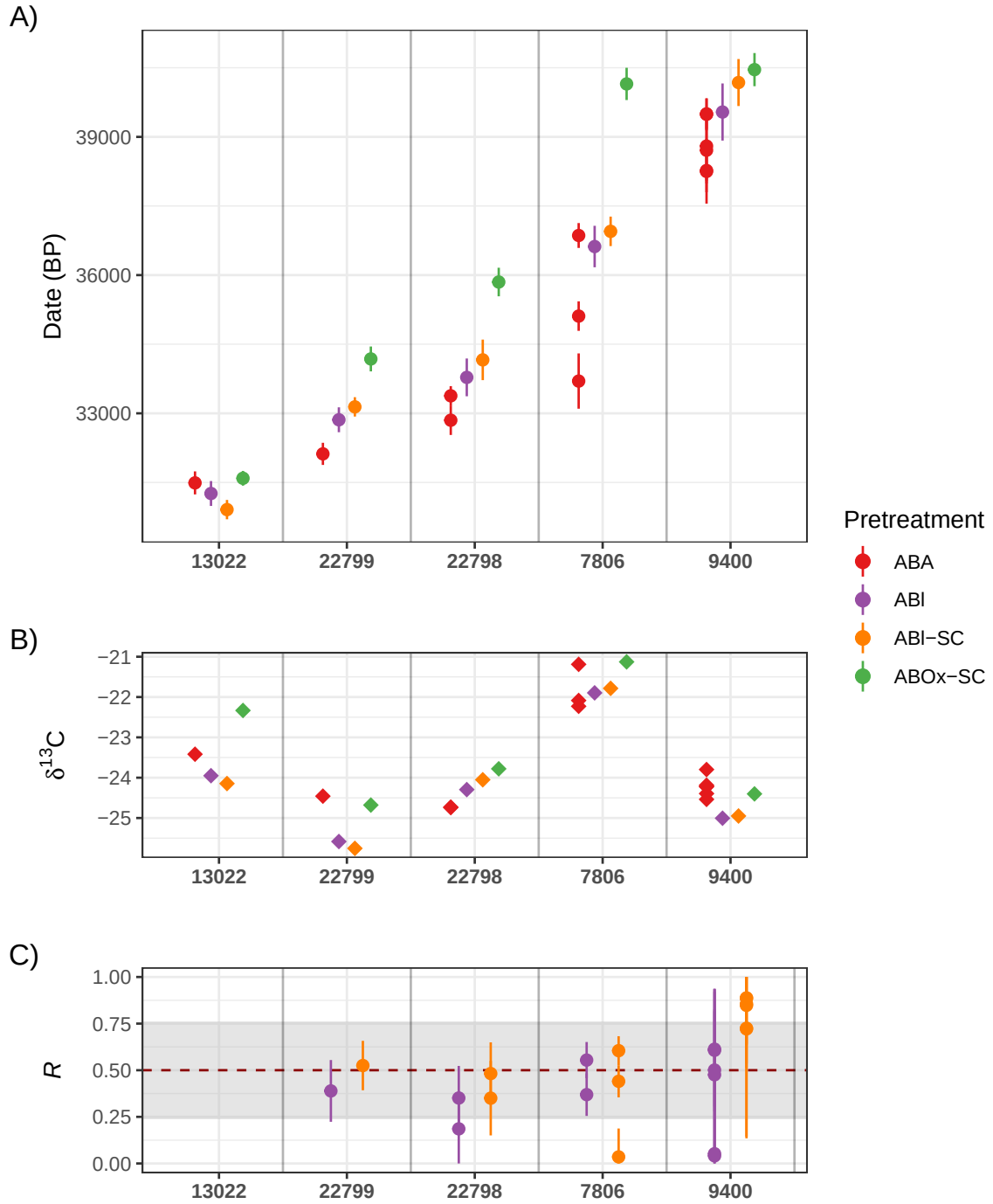


Figure 6.2: Summary of new ABI and ABI-SC radiocarbon determinations from Fumane, with ABA and ABOx-SC dates for comparison. Panel A shows radiocarbon ages for each sample for ABA, ABI, ABI-SC, and ABOx-SC pretreatment methods. Panel B shows stable carbon isotope ratios for the samples and methods. Panel C shows the proportional effect of ABI or ABI-SC over ABA, relative to the improvement over ABA by ABOx-SC (the higher the effect, the more contamination was removed, expressed as a proportion of the amount removed by ABOx-SC). Red dashed line and grey shading in Panel C show the mean and standard deviation across samples for R .

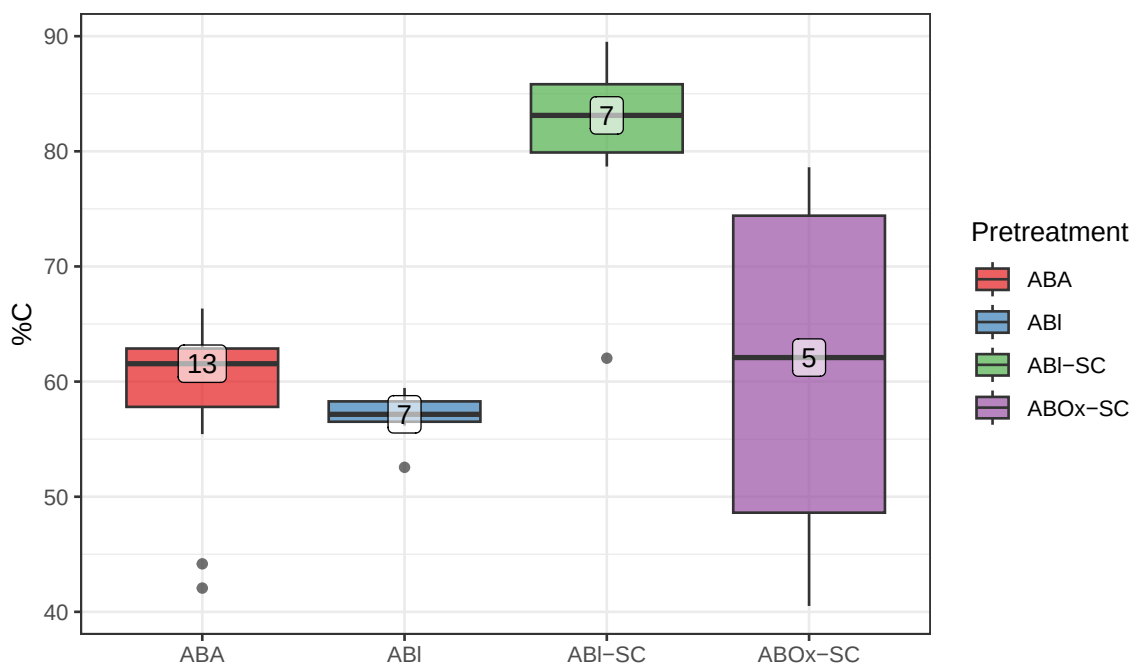


Figure 6.3: Boxplots showing %C measurements for ABA, ABI, ABI-SC, and ABOx-SC pretreatment methods. Numbers over boxplots indicate the number of samples summarised for each pretreatment.

For all ABI samples from Fumane and Renton Road, precombustion resulted in increased %C yields. %C increased from around 60% to above 75% in all cases from Fumane except 22798 (Figure 6.3, Table 6.3). The wider spread in %C yields for ABOx-SC, shown in Figure 6.3, is most likely explained by its distinct precombustion method and is discussed in Chapter 5.

With the exception of 22798, all archaeological samples showed a consistent pattern in $\delta^{13}\text{C}$ values: ABA and ABOx-SC pretreatments both showed enriched $\delta^{13}\text{C}$ values compared to ABI and ABI-SC. $\delta^{13}\text{C}$ values for 22798 are also in contrast to the pattern shown for other samples from Fumane, possibly related to the lower %C values for the sample, again due to the inclusion exogenous material such as sediment. Unlike the effect of precombustion on $\delta^{13}\text{C}$ values for ABA-treated samples (ABA vs. ABA-SC; Chapter 5), where precombustion enriched $\delta^{13}\text{C}$ was apparent for 6 out of 7 samples, $\delta^{13}\text{C}$ values for ABI and ABI-SC samples show no clear pattern, being similar before and after precombustion (Figure 6.2).

6.4 Discussion

In this chapter a novel chemical pretreatment was applied to charcoal samples, and its effect in contaminant removal tested. The goal of the experiment, following from results of Chapter 5, was to explore the effect of an alternative oxidation reagent in AOx-SC, and test its capacity to remove the exogenous fraction from charcoal before radiocarbon dating. The pretreatment method (ABl-SC) is identical to the AOx-SC protocol, except that potassium dichromate is replaced by sodium chlorite as an oxidation reagent. Charcoal samples from Fumane were used to test the method's effectiveness in contaminant removal, having previously been shown to produce significant differences between ABA and ABOx-SC dates. The experiment builds on a number of comparisons between ABA, ABOx-SC, and AOx-SC protocols (Douka et al., 2010; R. Wood et al., 2012), as well as Chapter 5's experiment exploring the effect of ABA with precombustion, ABA-SC. For samples from Fumane tested in this experiment, neither the experimental ABl-SC treatment or any of the aforementioned pretreatment methods were able to reproduce dates as old as those produced by ABOx-SC.

Unfortunately, there was not sufficient material available for this and Chapter 5's sampling requirements (~ 20 mg ABA excess for ABA-SC, *and* > 100 mg untreated material archived for ABl) to allow consistent and paired comparisons between this and the previous chapter's experiments; just three samples were dated by both ABA-SC and ABl-SC. Although some samples dated in Chapter 5 with lower excess may have survived chemical pretreatment under ABl-SC and could have been pursued, larger starting weights were preferred for the ABl method, given its experimental status and aggressive nature of the chemical treatment. These factors in combination meant that a dating greater number of samples by both methods was not feasible, meaning that the number of samples in the experiment is relatively small ($N = 5$). This and the broader effect of sample selection on experiments for the radiocarbon dating of charcoal are discussed in more detail in Chapter 7.

Results indicate that the ABl pretreatment method did not introduce modern contamination during pretreatment chemistry. Renton Road standards matched their expected radiocarbon ages, all being greater than 55 ky BP. It is relatively unsurprising that replacing potassium dichromate with sodium chlorite in the AOx-SC protocol does not introduce significant background contamination, given that sodium chlorite is a reagent frequently used at the ORAU for chemical pretreatment of various materials other than charcoal, and has not been shown to introduce contamination in these applications (Brock et al., 2010; Staff et al., 2014).

Chapter 5's results showed that precombustion is likely to be an essential aspect of the ABOx-SC protocol, accounting for around 40% of the method's improvement over ABA, but that precombustion alone does not account significant improvements in contaminant removal. The less-significant differences in radiocarbon dates between ABl and ABl-SC in comparison to the previous chapter's results (Figure 6.2), in combination with patterns shown in the ABl-treated samples' $\delta^{13}\text{C}$ values, suggest that either a strong base wash and/or aggressive oxidation is necessary to increase the lability of exogenous carbon in samples prior to precombustion (Tomiyaama et al., 2016). If this is the case, the significant differences between ABA and ABA-SC observed in the previous chapter are a strong suggestion that base washing contributes to increased contaminant lability. However, this is in contrast to the unpublished results of experiments in AOx-SC, where omission of the base step produces indistinguishable results. Considering these factors in parallel, it might be that either:

1. Distinct contaminant fractions are removed during wet chemistry stages and precombustion. Wet chemistry protocols do not interact with the precombustion process, and the lability of contaminants is not affected by their application.
2. Acidified potassium dichromate increases contaminant lability to the same degree as a strong base wash, and the labile components not removed during wet chemistry are removed during precombustion.

Further research into this problem, particularly exploring AOx or ABOx *without* stepped combustion is necessary, ideally under a comprehensive and iterative experiment design. This is discussed in more detail in Chapter 7.

Although results here highlight that high-strength oxidation and its relationship with precombustion are critical steps in removing contamination by ABOx-SC, some key uncertainties remain. There were multiple differences between the protocols used to produce ABOx-SC and ABI-SC dates:

- Oxidation reagents in ABOx-SC and ABI-SC are different (as per the experiment design)
- Following the AOx-SC protocol, ABI-SC does not include a base step
- Distinct precombustion methods were used to produce the ABOx-SC dates and ABI-SC dates, given that the precombustion procedure was changed at ORAU in 2012

The latter two differences preclude clear assessment of the effect of oxidation by sodium chlorite for paired ABOx-SC and ABI-SC radiocarbon determinations. In short, it could be any combination of these three changes that results in older dates for ABOx-SC.

In designing the experiment, the AOx-SC protocol was considered and adjusted for the purposes of the experiment, rather than ABOx-SC. This is because AOx-SC is routinely practiced at the ORAU, *not* ABOx-SC, and it was logical to adapt the former's protocol for the purposes of the experiment. In hindsight, a more comprehensive approach would have been to include a base wash prior to oxidation, allowing direct comparison between ABI-SC and ABOx-SC. However, this may have resulted sample failures: as noted above, justification for omission of the base step in AOx-SC is as yet unpublished, but internal experiments at the ORAU have suggested that the two methods are indistinguishable.

Four out of the five archaeological samples from Fumane show a similar pattern in their $\delta^{13}\text{C}$ measurements, with the exception being sample 22798. For the majority of samples, ABI-treated material is depleted in $\delta^{13}\text{C}$ in comparison to

ABA and ABOx-SC, both with and without precombustion. As considered above, and for ABA-SC in Chapter 5, this might suggest that distinct contaminant fractions with distinct $\delta^{13}\text{C}$ values are consistently removed by different pretreatment methods. However, unlike samples treated by ABA-SC, precombustion does not have a significant enrichment effect on ABl-treated material, except for a negligible difference seen for sample 22798. $\%C$ values are consistently much higher for precombusted ABl material, over 80% for some samples. This indicates that ABl in combination with precombustion results in a very carbon-rich product, but this does not necessarily mean that this carbon is endogenous to the sample. Given that ABl-SC is not able to produce dates as old as ABOx-SC, it is likely that a proportion of this carbon is exogenous contamination.

The combined suggestions from unpublished AOx-SC experiments and results presented here together raise some important questions for the chemical pretreatment of charcoal for radiocarbon dating. For example: exactly what contaminant fraction, and how much of it, is removed through oxidation by potassium dichromate? Could it be that high-strength HCl washing is also unnecessary in AOx-SC, and that aggressive oxidation followed by precombustion is sufficient for contaminant removal? And related, but broader: does acidified potassium dichromate really “do the same job” as an alkali wash in ABOx-SC? These questions should be answered by comprehensive and controlled experiments comparing the multiple differences between chemical pretreatments, and their various iterations. Again, the short-comings outlined above in this and Chapter 5’s experiments, and recommendations for the design of future experiments in radiocarbon dating of charcoal are discussed in more detail in Chapter 7.

6.5 Conclusion

A number of conclusions can be tentatively drawn from the results presented in this chapter:

1. Oxidation by sodium chlorite (ABl) is not as effective at removing contamination as the more aggressive oxidation by potassium dichromate in ABOx.
2. Precombusting ABl-treated material does not produce significantly older radiocarbon dates, suggesting that the wet chemistry of ABl does not increase the lability of contaminant fractions. Given precombustion after ABA produces significant differences to dates for ABA-treated material, as in Chapter 5, it could be that the base step of ABOx-SC and ABA-SC increases the lability of the contaminant fraction before precombustion, in contrast to the results of (unpublished) AOx-SC experiments.
3. Although not as effective as ABOx-SC at removing contamination, ABl-SC does not introduce modern contamination, as shown by ^{14}C ages >50 ky BP for the Renton Road samples.

7

Discussion of Experiments in Radiocarbon Dating

Abstract

Chapters 5 and 6 presented and discussed results of two separate but related experiments in radiocarbon dating, both focused on chemical pretreatment methods for charcoal prior to radiocarbon measurement. Samples from both experiments were from Grotta di Fumane, a cave site in northern Italy, where previous comparisons between charcoal pretreatment methods have been shown to produce significant differences in radiocarbon dates. In this chapter, results from both experiments are considered together, alongside the broader array of pretreatment methods for charcoal. The wider implications of the combined results, the common limitations of both experiments, and recommended directions for future research are discussed.

7.1 Summary of Experiments in Radiocarbon Dating of Charcoal

Chapters 5 and 6 presented two experiments in pretreatment chemistry prior to the radiocarbon dating of charcoal. Both experiments used alternative and adapted pretreatment methods to the two dominant protocols currently practiced at the ORAU: ABA and ABOx-SC. In many contexts, ABOx-SC has been shown to be more effective than ABA in removing exogenous carbon from charcoal samples, producing older dates than ABA at many Upper Palaeolithic sites that are often more consistent with stratigraphic ordering at their respective sites. Chemically speaking, it remains somewhat unclear exactly *why* ABOx-SC is so effective in contaminant removal. The two experiments aimed to explore and understand the difference between ABA and ABOx-SC in more detail, by adapting ABA and ABOx-SC (AOx-SC) protocols using samples that have previously produced significant differences in dates under two methods.

The first experiment (Chapter 5) showed that adding a precombustion step to the ABA protocol significantly *improved* contaminant removal over ABA. Indirectly, this provides some quantification of the effect of the stepped combustion stage of ABOx-SC. Results indicated that, on average, precombustion improved contamination removal over ABA by around 40%, expressed as a proportion of the inferred contaminant removal effect under ABOx-SC.

The second experiment (Chapter 6) explored an alternative, less aggressive and less toxic oxidation reagent in the ABOx-SC and AOx-SC protocols, replacing acidified potassium dichromate with sodium chlorite. Results showed that ABOx-SC without a base step and with a less aggressive oxidation reagent (ABl-SC) *worsened* contaminant removal. The less aggressive oxidation reagent was not as effective as acidified potassium dichromate under ABOx-SC, and without precombustion the method produced results that were largely indistinguishable from ABA in all but one case.

Results from samples dated in both experiments are shown in Figure 7.1, along with ABA and ABOx-SC radiocarbon determinations for comparison.

7.2 Samples Pretreated by Both ABA-SC and ABI-SC

Experimental treatments were applied to nine samples from Grotta di Fumane, but only three were dated in both experiments (22798, 7806, and 9400). For all three samples ABA gives the youngest date, implying the poorest contaminant removal of all pretreatment methods. Precombusting the ABA-treated material (ABA-SC) produced significantly older dates for these samples. Contaminant carbon loosely bound to the edges of aromatic structures of the charcoal was likely made more labile during the ABA wet chemistry, and removed by combustion and oxidation to CO₂ during the precombustion process (Tomiyaama et al., 2016). It is likely that the base step was a crucial aspect of this process in ABA.

Pretreatment by Acid-Bleach (ABI) shows further improvement over ABA and ABA-SC, although dates for all three samples pretreated by ABA-SC and ABI overlap at 1σ . Here, the implication is that a more aggressive acid wash and oxidation reagent act to remove contamination to the same degree as ABA-SC. Again, precombustion further improves contaminant removal following ABI (ABI vs. ABI-SC), but the difference is less pronounced than for ABA vs. ABA-SC. For all samples except 9400, ABOx-SC clearly outperforms all other pretreatment methods, but 9400's ABI-SC date is within error of its ABOx-SC date.

Stable carbon isotope results for samples dated in both experiments show consistent enrichment of $\delta^{13}\text{C}$ after precombustion (ABA vs. ABA-SC and ABI vs. ABI-SC), but no consistent pattern can be identified across all samples and pretreatment methods. Following precombustion using O₂ under ABA-SC and ABI-SC, %C values are consistently much higher than other methods, as expected.

From the pattern in dates shown Figure 7.1, a number of inferences can be made that may help to explain the effectiveness of ABOx-SC in comparison to all other pretreatment methods.

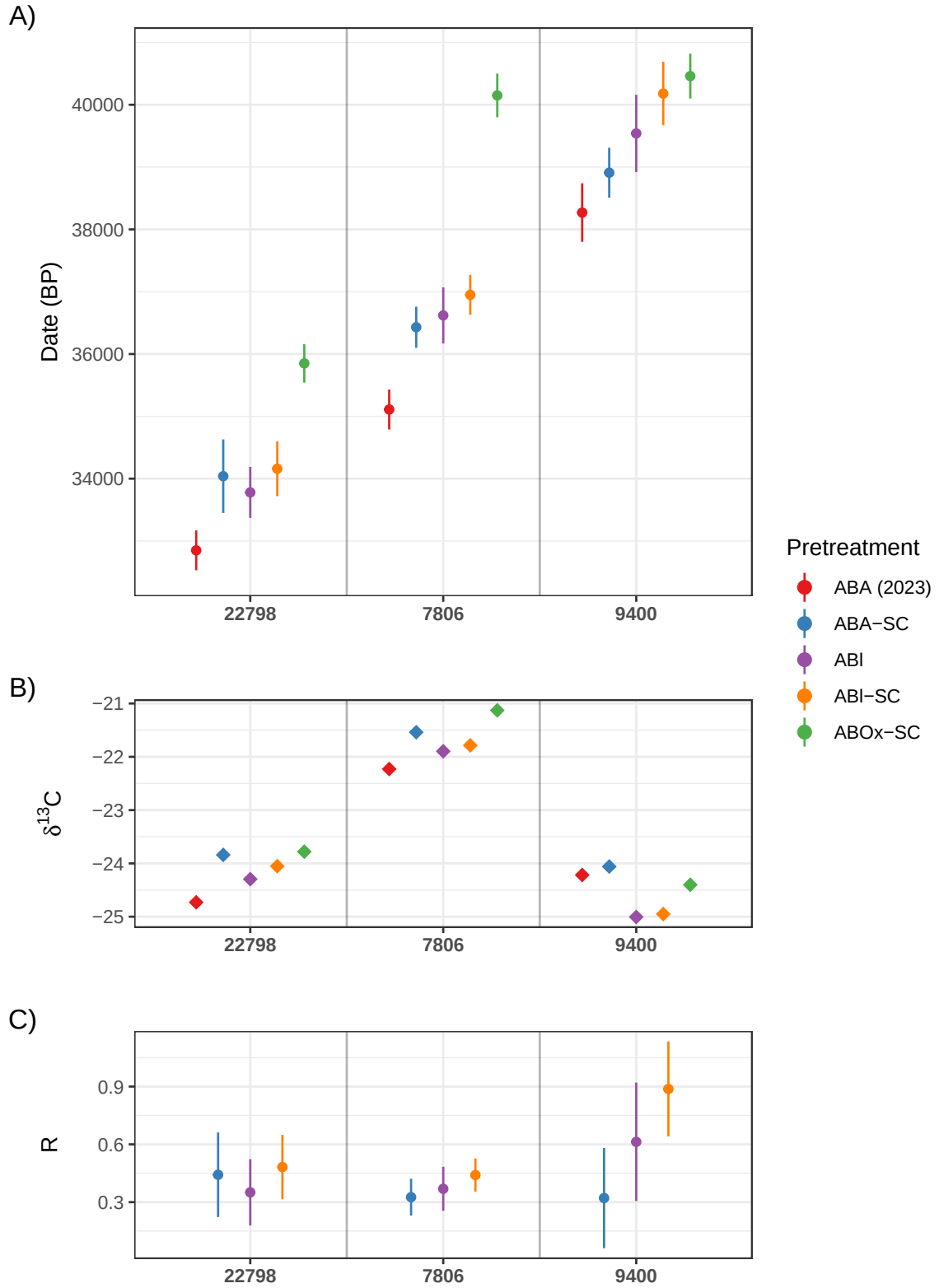


Figure 7.1: Three samples from Fumane dated by ABA-SC, ABI, and ABI-SC, compared to their ABA and ABOx-SC dates and $\delta^{13}\text{C}$ measurements. Panel A shows radiocarbon determinations for the samples, in ^{14}C years. Panel B shows $\delta^{13}\text{C}$ measurements. Panel C shows the improvement of each pretreatment method over ABA, expressed as a proportion of the improvement by ABOx-SC (Equation (5.1), Chapter 5).

First, it is clear that precombustion is an essential aspect of the ABOx-SC protocol, due to the removal of contaminants made more labile after wet chemistry, as demonstrated in Chapter 5. However, the proportional effect of precombustion is less dramatic following ABl than following ABA, and this may be due to interactions between wet chemical treatments and the precombustion process. Given that ABl does not use a base wash, it could be that the base step of ABA acts to increase the lability of the contaminant fraction by weakening contaminant bonds at the exterior of aromatic structures, allowing for precombustion to remove an additional proportion of the contaminant fraction more readily. This is supported by Bird et al. (1999)’s original introduction and discussion of the ABOx-SC method, as well as the broader theoretical foundation of chemical pretreatment of charcoal (Tomiyaama et al., 2016). However, as noted in Chapter 6, this explanation is in contrast with the implications of the (currently unpublished) experiments in AOx-SC, where the base step is omitted from the ABOx-SC protocol, but produces indistinguishable results. In order to fully understand and explore the interaction between base washes and precombustion, a comprehensive experimental design exploring ABOx-SC with a known contaminant fraction and signature is necessary, comparing results with and without both base washes and precombustion. Published results from the ORAU’s experiments in AOx-SC would also be informative.

Second, in all cases ABl-SC is the second-most effective pretreatment method, which suggests that both ABA and ABA-SC are improved by either: a) a stronger acid wash (increased from 1M to 6M HCl under ABl-SC); b) the addition of acidified sodium chlorite to the wet chemistry protocol; or c) both in combination. Again, the omission of a base step from ABl-SC precludes a direct comparison between ABl-SC and ABOx-SC, but given that ABl-SC cannot reproduce similar results to ABOx-SC in most cases, and given that for many contexts AOx-SC is indistinguishable from ABOx-SC, it is likely that the aggressive nature of oxidation by acidified potassium dichromate contributes significantly to contaminant removal under ABOx-SC.

Third, given that the pattern of results for varying pretreatment methods is not consistent across samples, it could also be that the various chemical protocols act to target and remove distinct contaminant types and fractions from the samples. For sample 9400, for example, ABl-SC and ABOx-SC produce similar results, but for 7806 and 22798 ABOx-SC clearly produces dates much older than any pretreatment method, including ABl-SC.

Whatever *type* of exogenous fraction present in 9400 can be removed by ABOx-SC and ABl-SC to the same degree, but the same cannot be said for 7806 or 22798. For the latter two, it might be that the source of contamination is distinct to that of 9400. ABA, ABl, and either with precombustion, are unable to target and remove this organic fraction effectively. Rather than ABOx-SC simply removing “more” contamination than other methods, it might be that each sample’s exogenous fraction is a distinct and variable array of organic contaminants, and that only ABOx-SC effectively targets this range of contaminants. In other words, ABOx-SC might be considered a “catch all” method, useful to remove a broad variety of contaminant types under the same protocol.

7.3 Implications of Experimental Results for Charcoal Pretreatment

All new radiocarbon measurements produced in the experiments in Chapters 5 and 6 are shown in Figure 7.2. ABOx-SC dates are also shown, and as described in previous chapters, these dates followed Bird et al. (1999)’s method for stepped combustion, using copper oxide for oxidation. A number of these samples have also been previously dated at the ORAU by other experimental methods, as well as many having multiple ABA dates. These are also shown in Figure 7.2. The additional experimental methods include plasma ashing and HCl wash (dated in early experiments with hypy), ABOx without stepped combustion, and ABOx-SC using O₂ precombustion.

One of the most striking observations made possible by collating and presenting the collection of dates for charcoal samples from Fumane is the wide spread of

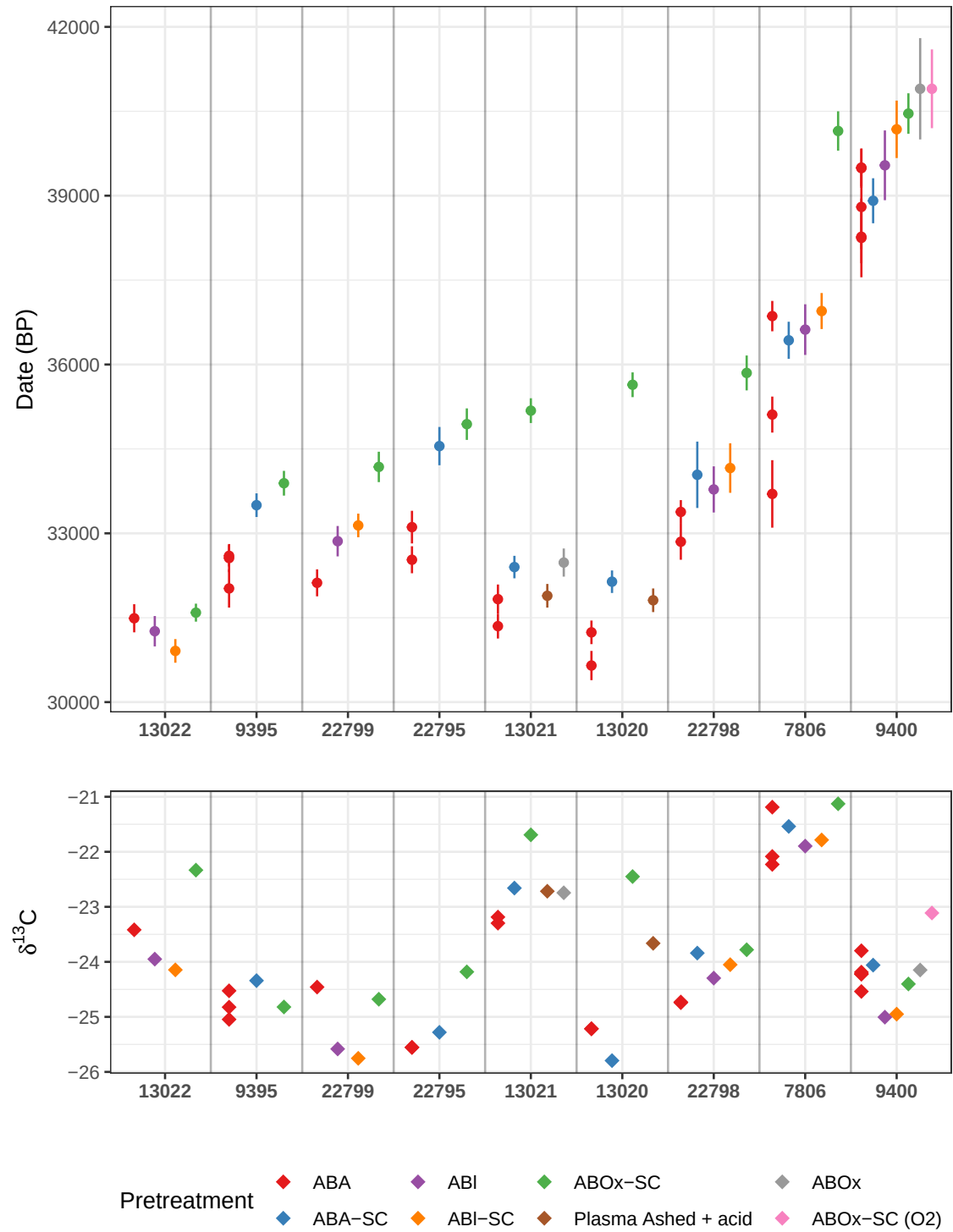


Figure 7.2: All available dates made at the ORAU for samples from Fumane, dated by ABA-SC, ABI, and ABI-SC methods (blue, purple, and orange, respectively). These include multiple ABA dates for each sample (red), ABOx without precombustion (grey), ABOx with CuO precombustion (green), ABOx with O₂ (pink) precombustion, and plasma ashed and HCl pretreatment (brown). Panel A shows radiocarbon determinations for each sample. Panel B shows $\delta^{13}\text{C}$ measurements for each sample. Considerable spread is particularly noticeable for ABA dates.

dates over the multiple methods used for each sample. In particular, significant variability is shown for dates produced by ABA; seven of the nine samples have multiple ABA dates, and all of them show significant differences in their radiocarbon determinations. Despite this variability, for most of these samples all their respective ABA dates are younger than any other method. However, variability in ABA dates raises the question: if there were multiple measurements for all other pretreatment methods considered, would they show similar variability?

Some variability in radiocarbon dates can be expected from repeated measurements, particularly for Palaeolithic samples, and it may be that the spread in ABA dates across all measurements is explained by varying levels of contaminant removal (or addition) during laboratory protocols, or variance and imprecision of the AMS. It is also possible that contamination levels vary within a single sample, and that variation in ABA dates is due to differences in contamination across different sampling locations, as discussed in Chapter 5.

If ABOx-SC can indeed act as a “catch all” pretreatment method, being able to remove very high proportions of exogenous carbon from multiple sources (as discussed above in Section 7.2), then *less* variability might be expected from multiple ABOx-SC measurements for the same sample. Unfortunately, no ABOx-SC measurements were repeated for samples from Fumane, likely due to the method’s laborious wet chemistry protocol. However, 9400 was dated by ABOx-SC, ABOx, and ABOx-SC with O₂ precombustion, which are all very similar methods, allowing for comparison for this sample.

The latter two methods show some difference from 9400’s ABOx-SC date, but these measurements are within error of each other, aligning with the idea that ABOx-SC might act to more consistently and aggressively remove organic contaminants of varying levels and from a variety of sources. Expanding on this point, an interesting experiment in the future would be to sample a single piece of contaminated charcoal multiple times and by multiple methods, in order to compare the spread of dates for each pretreatment protocol. From such an experiment,

some indication on the reliability and reproducibility of dates by each method could be given.

Sample 13021 has been dated by ABOx with and without stepped combustion, with the former producing a much older date than the latter. This supports the results from Chapter 5, which showed that stepped combustion significantly improves contaminant removal, and is an important aspect of the ABOx-SC protocol. However, sample 9400 was also dated by ABOx without stepped combustion, but both methods produce similar dates. Again, this suggests that the levels and types of contamination in charcoal samples at Fumane are variable, and that, currently, the only chemistry protocol able to remove such a range of contamination effectively and reliably is ABOx-SC.

Stable carbon isotope ratios show no consistent pattern across all samples. It is likely that the different pretreatment methods shown in Figure 7.2 act to target and remove distinct (or at least, variable) carbon fractions from charcoal samples, resulting in variability in $\delta^{13}\text{C}$ measurements. ABl and ABl-SC $\delta^{13}\text{C}$ values are typically lower than all other pretreatments, for example, while ABA-SC values are enriched compared to their ABA counterparts. Samples with multiple ABA measurements show variability in their $\delta^{13}\text{C}$ values, as for their radiocarbon measurements. Again, variation in contamination sources and levels, across and within samples, may contribute to and explain this variability.

Overall, the results presented here indicate that the effectiveness of the ABOx-SC pretreatment is multifaceted, and cannot be attributed to any one single aspect of the protocol. However, the results do not go as far as to explain exactly *how* ABOx-SC acts to remove contamination so effectively. Further experiments are necessary to understand this in more detail (see below, Section 7.5).

7.4 The Impact of Sample Selection on Experimental Pretreatment Methods and Using ABOx-SC as a Benchmark

In order to compare pretreatment methods robustly an appropriate set of samples is needed. Samples used in both experiments were initially selected on the basis that different charcoal pretreatment methods at Fumane have produced significantly different radiocarbon dates (Higham, Brock, et al., 2009). The sample selection process for both experiments was also limited by the amount of remaining material in storage at the ORAU. The experimental pretreatment designs required either sufficient ABA-treated material remaining after wet chemistry (Chapter 5), or enough untreated charcoal from the samples' original submission (Chapter 6). Due to these limitations, only a small number of total samples were available for either experiment, and even less could be included in both. In this section, the appropriateness of these samples for exploring experimental pretreatment methods are considered, outlining three common limitations between samples.

First, a major limitation common to both experiments is that the true age of the Fumane samples is not known. The impact of precombustion and alternative oxidation reagents in contaminant removal from samples from Fumane can only be expressed as a proportional effect, relative to ABOx-SC (R ; Equation (5.1)). Often, ancient ^{14}C -dead samples known to be contaminated are used as primary samples in experiments relating to pretreatment methods (Bird et al., 1999; Gillespie, 2019; Santos et al., 2001; Santos & Ormsby, 2013). In such cases, the effectiveness of the experimental pretreatment method can be measured by how *little* radiocarbon remains in the sample after pretreatment, or how close to background measurements the method achieves.

Previous studies assessed the accuracy of the ABOx-SC method by independent chronological markers, such as at Kostenki or Serino where the CI tephra horizon has been confidently identified (Douka et al., 2010; Higham, 2011; R. Wood et al., 2012). These experiments showed that dates by ABA consistently underestimated

the age of the CI-associated samples, whereas ABOx-SC did not. At Fumane, no known-age horizon or tephra layer has been found at the site that might help to confirm the accuracy of any dating method. The ABOx-SC dates at Fumane represent a minimum age for the samples; had any other experimental method produced an older date than ABOx-SC, it would have been considered more effective for contaminant removal, and the resulting date closer to the true age of the sample. As discussed in Chapters 5, this is because contamination by younger carbon is much more likely to affect a radiocarbon date than “old” carbon.

However, the reliability of the ABOx-SC dates at Fumane and their use as a benchmark in experiments is supported by a number of other factors:

1. ABOx-SC dates are in agreement with radiocarbon dates on ultra-filtered bone at the site.
2. Strong agreement was found for the resulting calibrated Bayesian model using ABOx-SC dates rather than ABA, and the model is altogether more consistent with age-depth patterns at the site.
3. Archaeological shifts at Fumane have been shown to align well typologically at other Italian Palaeolithic sites when dated by ABOx-SC, that have more securely dated sequences and where the CI has been found ([Higham, Brock, et al., 2009](#); [Higham et al., 2011](#)).

A second limitation of the Fumane samples is that the conditions of pyrolysis were not known. Experimental designs are possible that account for this, relating to [Bird & Ascough \(2012\)](#)’s “combustion continuum,” where a charcoal sample’s stability and interaction with contaminants within the deposition environment are related to temperature and oxygen levels during pyrolysis. The Maninjau charcoal appears to be very stable, but pyrolysis conditions for Fumane samples are completely unknown and are likely to vary between samples. Comparable experiments used ^{14}C -dead samples of fossilised wood to create charcoal under exactly controlled pyrolysis conditions to explore and compare charcoal pretreatments ([Bird et al.,](#)

2014, 2017; Orr et al., 2021; Tilston et al., 2016). Without knowing the conditions of pyrolysis, it is difficult to infer exactly what effect various pretreatment strategies have on contaminant removal, and whether one method is more effective for charcoals produced under certain conditions, for example. For archaeological charcoal from Palaeolithic contexts, however, it is very unlikely that comprehensive information on the conditions of pyrolysis can be known prior to radiocarbon measurement, although some broad inferences can be made (Constantine et al., 2021; Tintner et al., 2018). From this perspective, the “catch all” nature of ABOx-SC as inferred above has some clear advantages.

Lastly, and following from the previous point on the conditions of pyrolysis, the degree and nature of contamination in Fumane samples are also unknown. Given the variability in R across samples (e.g., Figure 5.5), it is unlikely that charcoal contamination throughout the Fumane sequence is consistent. Comparisons between samples show a consistent ordering of dates across methods, but their dispersion varies. Furthermore, their dispersion is not dependent on the age of the sample: for example, samples 7806 and 9400 are from close stratigraphic layers and show similar ABOx-SC ages, but ABA and ABA-SC appear to produce very different dates for each sample (Figure 7.2). Similarly, ABA-SC achieved very similar (statistically indistinguishable) results to ABOx-SC for 22795, but performed considerably worse for sample 13021, from the same stratigraphic layer. This might also suggest that the nature of contamination is variable for samples at Fumane, rather than simply the quantity: contaminants in sample 22795 appear more readily removed by ABA-SC than those in 13021, suggesting their composition and/or source may be distinct.

An experimental design in which the contaminant fraction of a sample is known and well-understood might be more appropriate to understand the effectiveness of the various pretreatment methods considered here. It is possible, for example, to deliberately contaminate exceptionally clean ^{14}C -dead charcoal with organic matter, spiking the sample with a ^{14}C signal that can be controlled and measured (e.g., leaf litter or limestone) (Bird et al., 2014, 2017; Orr et al., 2021; Tilston et al., 2016). The conditions of pyrolysis can be controlled in conjunction with known

contaminants: Bird et al. (2014) designed an innovative experiment on ^{14}C -dead charcoal produced at varying temperatures (300, 400, and 500 °C), exposing the samples to environmental contamination by leaf litter on a rainforest floor for up to three years. In a later experiment Bird et al. (2017) monitored and analysed the process of degradation, again with known contaminants. In such experiments, the exogenous fraction can be measured and quantified, in turn allowing precise quantification of its removal by various methods.

An expanded experiment would ideally use many more samples, of known age or with no endogenous ^{14}C , with a known and quantifiable contaminant fraction. However, charcoal from Fumane gives ample preliminary insight on the effectiveness of the experimental pretreatment methods presented here. Using known-age and known-contaminant samples is unlikely to change the main conclusion of these experiments: ABOx-SC is currently the most effective charcoal pretreatment method for charcoal.

7.5 Challenges and Recommendations for Future Research

The combined results of the experiments presented allow for:

1. Quantification of the effect of precombustion following ABA, expressed as a proportion of the contamination removed by ABOx-SC, estimated at around 40%
2. Confirmation that precombustion is an essential aspect of the ABOx-SC protocol
3. Indication that the use of acidified potassium dichromate under ABOx-SC is also likely to be another key aspect of the protocol, and that weaker oxidation reagents are not sufficient to remove contaminants to the same degree
4. Tentative inference that ABOx-SC can target and remove a greater *range* of contaminants, in higher proportions, than other pretreatment methods.

These experiments do not completely answer the question of *why* ABOx-SC is so much more effective than other methods. Moreover, these studies lead to key questions and recommendations for future research into the pretreatment of charcoal for radiocarbon dating, including the effectiveness of ABOx-SC. In this section, key recommendations are outlined and briefly expanded upon.

First, a comprehensive study into the effectiveness of ABOx-SC against other pretreatment methods (established, experimental, or otherwise), would ideally include a larger number of samples than were used in these experiments, and more than are typically used in other experimental pretreatment studies. The greater the number of samples, the greater the statistical confidence attributed to patterns in results, and in turn, the conclusions drawn from them. Many previous experiments in radiocarbon pretreatment for charcoal are hindered by their low sample sizes, which inherently limit the inferences that can be made from their results. Although radiocarbon dates are costly, experimental pretreatments can be expensive and at times wasteful, and acquiring sufficient amounts of material for experimental dating can be difficult, it is essential that future experiments employ larger sample sizes than has been preceded by previous experiments.

Second, measurements should be repeated several times for each sample. Results shown in Figure 7.2 indicate considerable spread for repeated ABA dates, which, as discussed above, raises the question of whether similar variability can be expected for other pretreatment methods. If, for example, repeated pretreatment and measurement by ABOx-SC gives more consistent results for one sample than other repeated methods, then ABOx-SC might be considered more accurate and reliable by way of its reproducibility, as well as its effective removal of contaminants. Repeating measurements for each sample also affords statistical certainty. However, as also noted in the previous point, increasing the number of measurements made in experimental pretreatments requires more material, laboratory hours, and ultimately, funding.

Third, the *type* of samples selected for experimental pretreatment should be better controlled and understood. Samples and their existing measurements from

Fumane were sufficient for preliminary testing of the experimental treatments designed and considered in Chapters 5 and 6. However, experiments in pretreatment methods for charcoal should ideally be supported by samples with the following attributes:

- **Being of known age (or, ^{14}C -dead)**

Using samples of known age allows for an exact quantification of both the level of contamination present in each sample, and the effectiveness of a given pretreatment method for its removal. Examples of such experiments include: Bird et al. (1999)'s introduction of the ABOx-SC method using ^{14}C -dead charcoal known to be ground-water contaminated; Douka et al. (2010)'s and R. Wood et al. (2012)'s use of CI-associated charcoals to verify sufficient contaminant removal by ABOx-SC; and Bird et al. (2014)'s innovative experimental design using a ^{14}C -dead charcoal "spiked" with a known contaminant. In these cases, the effectiveness of experimental protocols could be more readily assessed, as the true age of the samples is known or could be inferred from their context.

- **Having known conditions of pyrolysis**

Multiple studies have shown that the conditions of pyrolysis have an effect on the chemical stability of charcoal, and in turn, its susceptibility to contamination in the burial environment (Ascough et al., 2008; Ascough, Bird, Francis, Thornton, et al., 2011; Bird et al., 2017; Braadbaart et al., 2009; Dufourny et al., 2019; Tilston et al., 2016). As noted in Chapter 5, for all samples from Fumane the conditions of pyrolysis are unknown and are likely to vary between samples. Controlling for this using charcoal produced under exact and measured conditions could illuminate further the question of how ABOx-SC is so effective in contaminant removal in so many contexts and for so many samples, and how other methods might be improved.

- **Having known sources of contamination, of measurable and variable degrees**

Contamination in charcoal can be from a range of organic sources. Few studies have explored or compared the effectiveness of different pretreatment methods for varying contaminant types, despite key questions having been raised about the applicability of ABOx-SC to samples deposited in clay-rich soils, for example (Alex et al., 2017; Rebollo et al., 2011). If one pretreatment method is shown to be more effective than another for contaminants from specific sources, it might be possible to design context-specific pretreatment methods that target and remove the most likely contaminant types, according to the depositional environment. Similarly for samples of known age, controlling for different degrees of contamination will also allow for an assessment of the performance of various pretreatment methods, and give some quantitative indication of their effectiveness for contaminant removal.

- **Being from a range of geological contexts.**

If environmental and archaeological samples are to be used in experiments (i.e., not laboratory pyrolysed and not ^{14}C -dead samples that have been deliberately contaminated), they should ideally be from a range of geological contexts that encompass or encourage a wide range of contaminant types. This approach is less exact than that of the previous point (using samples of deliberate and measurable contamination), but might be a more useful approach in terms of the “real world” application of innovative pretreatment methods. Rarely in radiocarbon dating is the true age or level of contamination known before pretreatment, and developing a more comprehensive suite of protocols for a variety of depositional contexts would be a useful endeavour.

Fourth, the differences between pretreatment methods should ideally be incrementally explored and quantified, with all variables and differences between methods considered and tested in parallel. For the experiments presented here,

adapted ABA and ABOx-SC protocols were designed and tested, after highlighting three key differences between the methods. Only two of these differences were explored in experiments: applying a precombustion step to ABA (ABA-SC); and replacing potassium dichromate with a less aggressive oxidation reagent under AOx-SC. However, the effect of the third difference between ABA and ABOx-SC (increased acid and base concentrations) remains unquantified and unexplored. A more comprehensive experiment might gradually increase these concentrations, alongside multiple alternative oxidation reagents, with and without precombustion, for example.

Incrementally and iteratively adjusting protocols in a logical design would help to avoid ambiguity in future experiments, and to fully understand the differences between ABA and ABOx-SC. [Tomiyama et al. \(2016\)](#) analyse and date multiple fractions for the three environmental samples used in their study, for example. However, controlling for all of these variables quickly expands such an experiment to a very large study, requiring multiple sites, sample types, and contaminant types, as well as a comprehensive suite of methods that incrementally and iteratively measure the many methodological differences between ABA and ABOx-SC.

Fifth, and following from the previous point, a comprehensive experiment should include multiple chemical and elemental analyses of the products of the various stages of pretreatment steps, alongside their final radiocarbon measurements. Incrementally adjusting protocols (e.g. increasing acid and base concentrations) and conducting such analyses at each increment would afford greater understanding of *what* is being removed (or left behind) at stages of the protocols, as well as *how* this happens. FTIR or Raman spectroscopy are two relatively simple analyses without excessive cost or labour. By comparing FTIR spectra of charcoal products before and after precombustion, some indication of *what* components are being removed so readily during this step, to result in significantly increased proportions of carbon in a sample (%C), for example.

Finally, a straightforward experiment that would further illuminate the effect of stepped combustion in ABOx-SC would be to chemically analyse and date excess

ABOx-treated material (*without* stepped combustion) stored at the ORAU, and compare results to existing measurements for ABOx *with* stepped combustion. Figure 7.2 shows two ABOx measurements with and without stepped combustion, 13021 and 9400. For 13021, the effect of stepped combustion is clearly significant, suggesting that a large proportion of contaminants were removed from the sample during this step. For 9400, by contrast, stepped combustion (and precombustion with O₂) has little effect on radiocarbon determinations, suggesting that the majority of contaminants (that other pretreatment methods failed to remove) were removed during ABOx wet chemistry. These two dates alone show that precombustion does not always have significant effects on ABOx-treated material, highlighting the possibility that distinct contaminant types are targeted and removed by the various stages of the ABOx-SC protocol. Again, combining this experiment with more detailed chemical analysis (FTIR or Raman spectroscopy) as described above could also be informative.

8

An Updated Radiocarbon Chronology for the Upper Palaeolithic Sequence at Grotta Paglicci, Italy

Abstract

This chapter presents an updated radiocarbon chronology for the Upper Palaeolithic stratigraphic sequence at the site of Grotta Paglicci (Puglia, southern Italy). Grotta Paglicci has a rich archaeological record, providing one of the most complete Upper Palaeolithic sequences in Europe, from the early Aurignacian through to the late Epigravettian. The updated chronology is based on a series of new accelerator mass spectrometry (AMS) dates on charcoal samples recovered from Aurignacian to Gravettian layers (~ 40 - 20 ky calBP). The majority of samples were pretreated using AOX-SC or ABOx-SC methods, broadly considered the most effective pretreatments for charcoal samples older than 25 ky BP. Radiocarbon dates were calibrated using IntCal20, the most recent version of the northern hemisphere terrestrial radiocarbon calibration curve ([Reimer et al., 2020](#)).

Cultural transitions at Grotta Paglicci were further constrained by Bayesian modelling. Two Bayesian models were built: one using all available dates; and a second using only the new dates produced for this study. The timing of cul-

tural shifts at Grotta Paglicci, based on the two Bayesian models, are compared to relevant palaeoenvironmental records and discussed in the context of large-scale environmental change across Europe during MIS 3 and MIS 2. Given the excellent preservation of charcoal at the site and the now comprehensive set of radiocarbon measurements made throughout the Upper Palaeolithic sequence, the updated chronology for Grotta Paglicci represents one of the most up-to-date and robust chronologies of Upper Palaeolithic cultural change in Europe before the Last Glacial Maximum.

As for many other recently published radiocarbon chronologies for this period, the Paglicci chronology presents an opportunity to explore the relationship between cultural and environmental change during MIS 3 in Europe. However, comparisons to the Greenland ice core record and pollen records from Tenaghi Philippon, Monticchio, and Ioannina show that, as detailed in Chapter 2 and Chapter 3, the sampling resolution of many of these records, along with the uncertainties associated with their chronologies, preclude any assessment of environmental causality for centennial-to-millennial-scale archaeological changes. These comparisons show that framing the emergence of the Aurignacian or Gravettian as direct behavioural responses to stadial or interstadial conditions is problematic. By considering the *duration* and *persistence* of these industries alongside environmental records, rather than their emergence, the relationship between these processes can be better understood. In short, if the question cannot be answered by the data available, the question should be reframed to one which can.

8.1 Introduction

The European Upper Palaeolithic holds some of the most studied and well understood records of technological and cultural change in Palaeolithic archaeology. The Aurignacian and Gravettian technocomplexes have been a central focus of this research, together spanning a wide chronological window (~ 20 kyr) with a broad geographic distribution. Both the Aurignacian and Gravettian are defined by and associated with numerous technological innovations and behavioural shifts identifiable in the archaeological record (Bar-Yosef, 2002; Hoffercker, 2005, 2009). One of the key questions for this broad period is the extent to which these archaeological changes can be causally linked to large-scale environmental variability observed in palaeoclimate records (Badino et al., 2020; Banks et al., 2013; Reynolds & Green, 2019). Radiocarbon dates and robust chronologies are critical to understanding this question, as well as the spatio-temporal character of human innovation, interaction, and movement (Higham, 2011; Mellars, 2006).

The Aurignacian has been fundamental to understanding the arrival and dispersal of anatomically modern humans (*Homo sapiens*, AMH) in Europe, along with the dynamics of technological innovation and interaction during the early Upper Palaeolithic. Whilst authorship of earlier “transitional” industries has been debated, such as the Initial Upper Palaeolithic, the Chatelperronian, the Uluzzian, or the Bohunian (Benazzi et al., 2011; Douka et al., 2014; Fewlass et al., 2020; Goring-Morris & Belfer-Cohen, 2020; Higham et al., 2010; S. L. Kuhn, 2019; S. L. Kuhn & Zwyns, 2014; Marciani et al., 2020; Peresani et al., 2016; Villa et al., 2019; Welker et al., 2016; Zilhão et al., 2015), the producers of the Aurignacian are widely accepted to be AMH, with the industry appearing in Europe around the time of or relatively soon after the disappearance of Neanderthal populations, around 43 - 42 ka calBP (Barshay-Szmidt et al., 2018; Douka et al., 2012; Higham et al., 2012, 2014; Nigst et al., 2014; R. Wood et al., 2014). The arrival of the Aurignacian tool kit at sites across Eurasia (Proto, Ancient, Early, or otherwise) represents the irreversible shift from Middle Palaeolithic to Upper Palaeolithic industries, and

by extension, the unequivocal and established presence of AMH in a given region. Edge-retouched Aurignacian blades and split-based bone or antler points are typical examples from the “Classical” Aurignacian tool kit (Mellars, 2006), while the industry is also associated with wider behavioural and technological changes of the Upper Palaeolithic. Increasing frequencies of art, musical instruments, personal objects, and smaller and more specialised stone tools are found in Aurignacian contexts, in comparison to earlier periods (Hoffecker, 2005).

The Gravettian represents a similarly significant and well-documented technological and cultural shift in the Upper Palaeolithic, showing a marked change toward microlithic and backed technologies of the mid-Upper Palaeolithic. The Gravettian is a diverse and complex archaeological category (Reynolds et al., 2015), but can be broadly characterised by the production of long, narrow and straight blades and bladelets, the development of a “modular” and more portable tool kit in comparison to the Aurignacian, and changes in raw material selection for tool production, most notably increased frequencies of osseous tools (Borgia et al., 2011; Reynolds & Green, 2019; Taller & Conard, 2022). As seen for the Aurignacian, a number of additional innovations emerge alongside changes in tool production during the Gravettian, including lamps, clothing, eyed needles, food storage, and fired clay objects (Hoffecker, 2009). Other changes during the period appear more symbolic, with higher frequencies of burials, changes in funerary practices, and the production of small portable objects such as the well-known Venus figurines. Although the exact origins of the Gravettian are unknown, it is considered a “pan-European” phenomenon: its modular tool kit appears to have spread relatively quickly after its design and is identified at sites across Europe, from Portugal to Western Russia (Bicho et al., 2015; Kozowski, 2015; Reynolds et al., 2015; Svoboda, 2007).

A recurring question in the study of both the Aurignacian and Gravettian industries concerns the association between their numerous technological changes and variability in environmental conditions during Marine Isotope Stage (MIS) 3

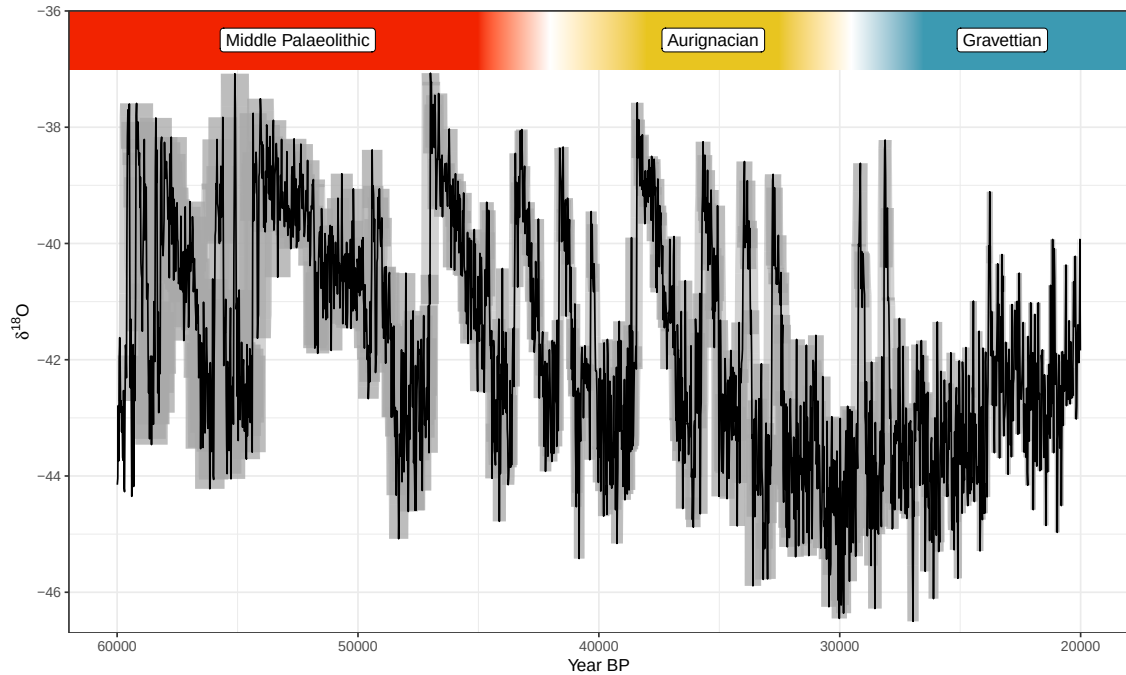


Figure 8.1: $\delta^{18}\text{O}$ record from NGRIP (black) with its chronological uncertainty in grey (Rasmussen et al., 2014), and approximate timings for the emergence and duration of Middle and Upper (Aurignacian and Gravettian) Palaeolithic industries in Europe (Douka et al., 2020; Higham et al., 2014).

and 2 (Figure 8.1). Upper Palaeolithic archaeological sequences are typically discussed with reference to records of environmental change, such as ice cores, pollen records, or speleothems. Climatic periods of varying character have been employed as archaeological “push or pull” factors to explain changes in the archaeological record. Many Upper Palaeolithic innovations have been associated with, compared to, and explained by climate stability, amelioration, or deterioration (Andel, 2005; Badino et al., 2020; Bataille & Conard, 2018; Falcucci et al., 2020; Marín-Arroyo et al., 2023; Nigst et al., 2014).

“Eco-cultural” niche modelling has been a focus of the Aurignacian, for example (Andel, 2005; Banks et al., 2013; Hauck et al., 2018; Riel-Salvatore & Negrino, 2018; I. Schmidt & Zimmermann, 2019), where specific behavioural changes are explained by changing adaptive niches, modelled statistically as responses to environmental or topographic variables. Similarly, innovations during the Gravettian have frequently been discussed as representing behavioural responses to colder and drier conditions that persisted during stadial periods (Andel, 2005; Arrizabalaga et al., 2021;

Beresford-Jones et al., 2011; Haws et al., 2021; Maier, 2017; Taller & Conard, 2022).

As discussed in Chapter 2 and Chapter 3, environmental records for MIS 3 are complex and highly variable in their characteristics. The Greenland ice core record, one of the most frequently referenced records of palaeoenvironmental change in the archaeological literature of the Palaeolithic, shows pronounced *variability* over the Late Pleistocene, during MIS 3 in particular. Abrupt temperature oscillations during DO events (see Chapter 2) and transitions between interstadial and stadial periods during MIS 3 and MIS 2 occurred over periods as short as several decades (Dansgaard et al., 1993; Kindler et al., 2014; Rasmussen et al., 2014). The durations of stadials and interstadials, as well as their overall character, vary considerably throughout MIS 3, whilst comprehensive explanations for the dramatic changes and the causes of transitions remain illusive (Izumi et al., 2023; Muglia et al., 2023).

On the Italian Peninsula, one of the most important sites for understanding the emergence of the Aurignacian and the subsequent transition to the Gravettian is Grotta Paglicci (Paglicci; Palma di Cesnola, 2004). Current radiocarbon chronologies suggest the site's Gravettian levels are among the earliest in Europe, leading some to argue that the industry may have originated in the region, developed as a response to environmental conditions (Palma di Cesnola, 2006). A large number of radiocarbon dates exist for the site, but most of these measurements were made over 40 years ago with undocumented pretreatment methods.

This chapter presents a new series of accelerator mass spectrometry (AMS) dates for charcoal recovered from Aurignacian and Gravettian layers at Paglicci, using ABOx-SC, AOx-SC, and ABA pretreatment methods, calibrated using IntCal20 (Brock et al., 2010; Reimer et al., 2020). Using the updated chronology from Paglicci, the relationship between cultural innovation and environmental change is considered, and the limitations of the approach discussed.

Section 8.2 describes Paglicci and its archaeological sequence, followed by an assessment of existing radiocarbon measurements and chronological information at the site. In Section 8.3, radiocarbon and Bayesian methods are outlined, with

results presented in Section 8.4. Finally, the findings of this study and comparisons between environmental and archaeological records on the Italian peninsula are discussed in Section 8.5, along with recommendations for future research.

8.2 Grotta Paglicci

Grotta Paglicci (Paglicci) is a cave site located in Southern Italy, on the south-western slopes of the Gargano Promontory of the Puglia region (Figure 8.2; 41° 39' 14.5" N, 15° 36' 54.5" E). The site lies around 140 m above sea level, above the Tavoliere Plain (*Tavoliere delle Puglie*) to the south-west, below the rocky peaks of the Gargano Promontory to the north-east, around 30 km to the west of the Adriatic Sea. The Gargano Promontory is complex and rocky in its topography and today is characterised by a Mediterranean climate, while the Tavoliere Plain is flat, arable, and extensively cultivated. During periods of the Late Pleistocene the Tavoliere Plain was likely much colder and drier than today, being dominated by wooded steppe, and this might be extended to the Gargano Promontory (Borgia et al., 2016; Lippi et al., 2015).

Paglicci has a long history of excavation since the 1960s. Excavations were first directed by the Museo de Civico di Storia Naturale di Verona, but since 1971 have been managed by the University of Siena in collaboration with the Office for the Archaeological Heritage of Puglia (Boschin et al., 2020; Ronchitelli et al., 2015).

8.2.1 Archaeological Sequence

Two areas of the site have been excavated: an external rock shelter; and a deep trench in the main cave system. Sediments from the external rock shelter have yielded Acheulean and Middle Palaeolithic (late Mousterian) stone tools, among evidence of hyena activity and ungulate remains (Crezzini et al., 2016). The rock shelter appears to have gradually reduced in size through Acheulean and Middle Palaeolithic phases, due to repeated roof collapse. No later deposits were present above Mousterian layers in the external shelter, likely due to the roof's final collapse during this period (Boscato & Ronchitelli, 2006; Crezzini et al., 2016).

Radiometric dates have not been attempted for the external rock shelter sequence. Dating by biochronology places its uppermost layers in the Middle to Late Pleistocene, based on the presence of *Iberomys brecciensis*, a small rodent that

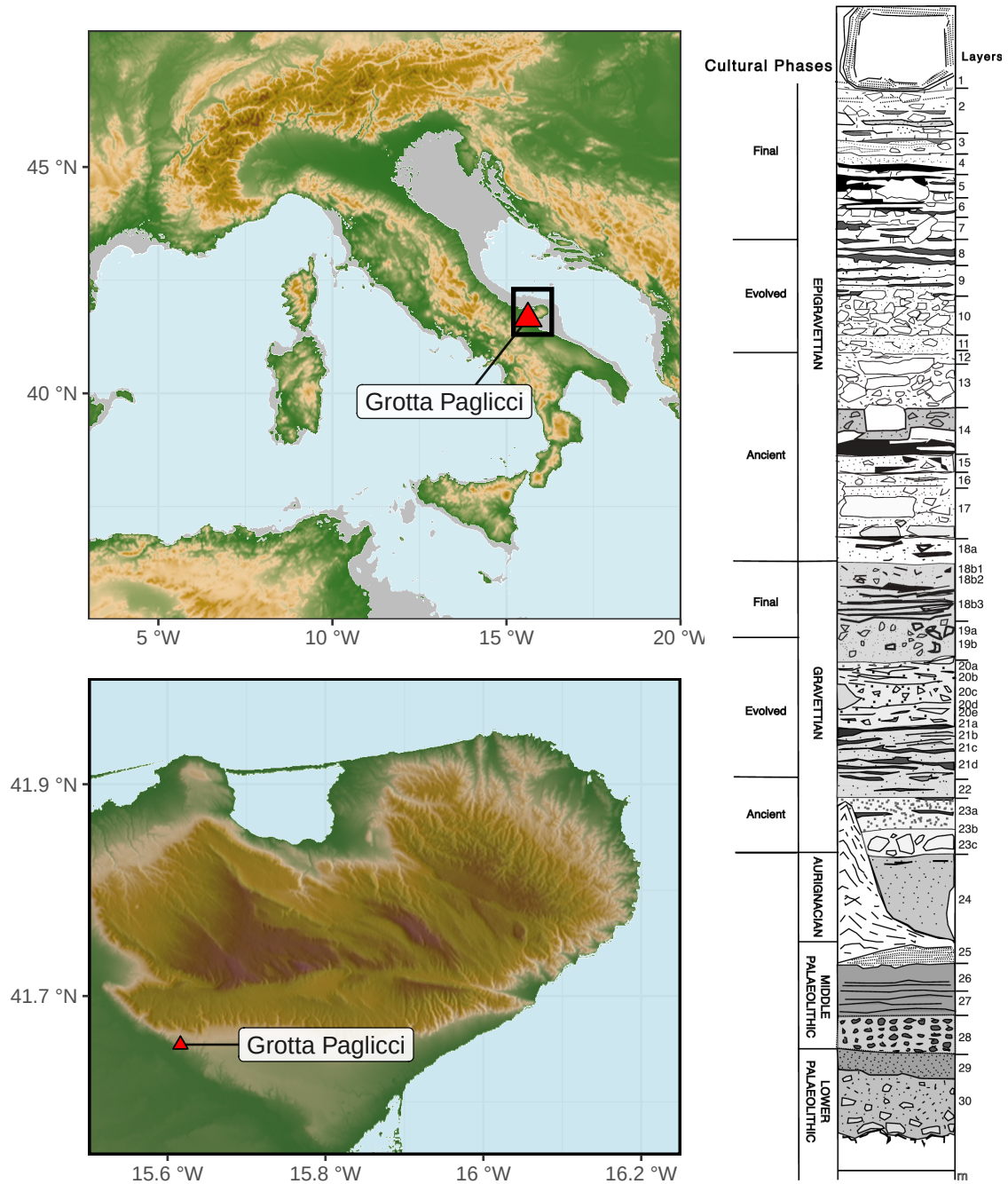


Figure 8.2: Location of Grotta Paglicci in Italy (top left), its location on the Gargano Promontory (bottom left), and the stratigraphic diagram of layers of the interior cave sequence and major archaeological phases at the site (right, redrawn from Boschini et al., 2018).

became extinct before the last glacial period (Crezzini et al., 2016; López-García et al., 2015). Ungulate assemblages from the external sequence are dominated by *Cervus elaphus* and *Dama dama* (red deer and fallow deer, respectively), suggesting forested environments in the region before c. 125 kyr BP, in contrast to periods of the Late Pleistocene (Arobba et al., 2004; Crezzini et al., 2016).

The interior cave sequence comprises 30 stratigraphic levels identified across 8 metres of excavated sediment (Figure 8.2). From these levels, archaeological material has been recovered relating to the Lower and Middle Palaeolithic (Layers 30 - 26), through the Aurignacian, Gravettian, and Epigravettian of the Upper Palaeolithic (Layers 24 - 1; Figure 8.2). Lower and Middle Palaeolithic layers are separated from the Upper Palaeolithic by a sterile hiatus in Layer 25, a thick deposit of red clay that has yielded no archaeological material. This study concerns Layers 24 to 18, relating to Aurignacian and Gravettian industries of the Upper Palaeolithic.

The Aurignacian is only found in Layer 24 at Paglicci. Layer 24 is comprised of a sandy-silt sediment, approximately 1 meter thick and divided into sub-layers by sediment colour: BII, BI, A4, A3, A2, A1, and A0 (Palma di Cesnola, 2006). Layer 24 is also divided archaeologically into three industrial phases: Ancient Aurignacian (sub-Layers BII to BI); Middle Aurignacian (A4 to A2); and Upper Aurignacian (A1 to A0). The archaeological subdivisions of the Aurignacian at Paglicci are defined by changes in the proportions of tool types, with middle and upper phases having yielded larger quantities of stone tools, and larger proportions of retouched bladelets (*Famile des Retouches différenciés*; Palma di Cesnola 2006: 356). Throughout all three archaeological phases, populations at Paglicci produced marginal backed and Dufour bladelets. The marginal backed bladelets from Layer 24A1 are somewhat unique and represent a “peculiar aspect” of the Upper Aurignacian (Borgia et al., 2016: 161), being a distinct tool type found only at Paglicci, but possibly linked to other industries in Europe (Stepanchuk & Vietrov, 2021).

From the Aurignacian levels at Paglicci, along with chronologies from corresponding layers at other Aurignacian sites in the north (Riparo Mochi and Fumane) and south of the Italian Peninsula (Castelcivita, Serino, Grotta della Cala, and Grotta della Calanca), [Palma di Cesnola \(2006\)](#) argues that the Aurignacian entered the region by the movement of human populations from the north between GS 11 and the Arcy interstadial (possibly corresponding to GI 5.1), between 42 - 32 ky BP. According to [Palma di Cesnola \(2006\)](#), the incoming Aurignacian industry replaced the Uluzzian in the south of Italy, before populations became “settled” in regions such as Gargano Promontory (*“les groupes aurignaciens à dos marginal du Sud de la Péninsule se seraient installés”*; Palma di Cesnola 2006: 366). This goes some way to explain the preceding hiatus (Layer 25) and absence of Proto-Aurignacian tool types at Paglicci. [Palma di Cesnola \(2006\)](#) argues that these populations developed novel and specialised tool types, represented by the earliest Gravettian at Paglicci and other Upper Palaeolithic sites in Italy, such as Rio Secco and Grotta di Fumane ([Falcucci & Peresani, 2019](#); [Peresani et al., 2014](#); [Talamo et al., 2014](#)).

The earliest Gravettian at Paglicci is found in Layer 23, a yellow-brown sandy deposit about 70 cm thick, subdivided into horizons A, B, and C, with horizon C being further divided to CI and CII. Archaeologically, Layers 23 and 22 at Paglicci are defined as Early (Ancient) Gravettian, by the appearance and abundance of backed projectile points that have been suggested to reflect the development of new hunting strategies ([Palma di Cesnola, 2006](#); [Wierer, 2013](#)). In Layer 23C the frequency of human occupation at Paglicci appears more sporadic, compared to an intensification of activity in later Gravettian layers. Layers 23 and 22 show little variation in their lithic assemblages. Layer 23A yielded a pestle grinding tool, with residue analysis suggesting complex, multistep plant processing ([Lippi et al., 2015](#)). Layer 22 is a 20 cm deposit subdivided from F to A, with numerous hearths dug at depths of approximately 20-30 cm ([Palma di Cesnola, 2006](#)). Overlying Evolved Gravettian layers (Layers 21 - 19B) are also “intensely anthropized,” with two burials found in Layer 21 ([Ronchitelli et al., 2015](#): 424). Wall paintings at Paglicci

are attributed to the Final Gravettian or Early Epigravettian, corresponding to Layer 18 and above (Zorzi, 1964).

Aside from the Proto Aurignacian, the sequence at Paglicci represents a complete and continuous Upper Palaeolithic sequence, from the Aurignacian to Epigravettian. Some “local” nuances and variability are apparent across horizons. To understand whether these nuances represent innovation or “origins” of industries, a robust and precise chronology for the sequence is necessary. In the next section, a review of existing radiocarbon measurements is provided.

8.2.2 Previous Radiocarbon Measurements and the 23C2 Tephra

Previously published radiocarbon data for Grotta Paglicci are presented in Table 8.1, and calibrated dates are plotted in Figure 8.3. These measurements were collated by Berto et al. (2017) and listed in the Radiocarbon Palaeolithic Europe Database (Vermeersch, 2020). According to Vermeersch (2020), 7 of the 53 available radiocarbon dates at Paglicci were measured by AMS, with the remainder by “conventional” ^{14}C measurement. These dates were made at the Florence, Groningen, Lyon, Rome, Utrecht, and Oxford laboratories (Vermeersch, 2020). Together, they suggest that Layers 24 to 18 pre-date the Last Glacial Maximum (LGM), corresponding to MIS 3 and the beginning of MIS 2. As yet, Bayesian methods have not been used to further constrain radiocarbon dates from Paglicci.

Detailed records for these radiocarbon measurements could only be found for dates measured at the Florence, Lyon, and Oxford laboratories (Azzi et al., 1974, 1977; Evin et al., 1979). Relevant documentation is not readily available for measurements made at the Groningen, Utrecht, and Rome laboratories. For many dates at Paglicci, the pretreatment chemistry, year and method of ^{14}C measurement, and even material dated, are either unknown or could not be confirmed by laboratory documentation (Table 8.1).

At least one error exists in the literature and databases of existing radiocarbon dates from Paglicci. One of the Florence dates for Layer 21C is given the lab code

Table 8.1: Previous radiocarbon determinations for samples from Grotta Paglicci.

Layer	Sublayer	Method	Lab Code	Material	Year Dated	Pretreatment	Reference	Date Excavated	Date	±
3	A	Gas counting	F-94	Charcoal	1976	ABA	Azzi 1977	1972-1975	11440	180
4	AC	Gas counting	F-95	Charcoal	1976	ABA	Azzi 1977	1972-1975	11950	190
5	BC	Gas counting	F-96	Charcoal	1976	ABA	Azzi 1977	1972-1975	13590	200
6		Gas counting	F-64	Charcoal	1976	ABA	Azzi 1977	1972-1975	14270	230
7	C	Liquid scintillation	Ly-1628	Bone	1977-1978	HCl + H ₂ O	Evin 1979	1973	13720	870
7		Gas counting	F-65	Charcoal	1976	ABA	Azzi 1977	1972-1975	14820	210
8		Gas counting	F-66	Charcoal	1976	ABA	Azzi 1977	1972-1975	15460	220
9		Gas counting	F-67	Charcoal	1976	ABA	Azzi 1977	1972-1975	15270	220
10		Gas counting	F-68	Charcoal	1976	ABA	Azzi 1977	1972-1975	15320	250
12	A	Con ¹⁴ C	GrN-14316						15950	350
12	C	Con ¹⁴ C	GrN-14317						15730	330
13	A	Con ¹⁴ C	GrN-14318						15480	150
13	B	Con ¹⁴ C	GrN-14319						16310	350
13	B	Con ¹⁴ C	GRN-14321						16030	190
13	C	Con ¹⁴ C	GRN-14320						15990	160
14	B	AMS	UtC				Mussi 2006		15600	200
14	B	Con ¹⁴ C	GrN-14322						15930	200
15	A	Con ¹⁴ C	GrN-14323						15750	160
15	B	Con ¹⁴ C	GrN-14324						16260	160
15		Con ¹⁴ C	UtC						17100	300
16	A3-1	AMS	UtC						17200	300
16	A3-1	Con ¹⁴ C	GrN-14325						16690	150
16	B3	Con ¹⁴ C	GrN-14326						16450	190
16	B3B	Con ¹⁴ C	GrN-14870						16970	130
16	B5	Con ¹⁴ C	GrN-14871						16750	150
16	B7	Con ¹⁴ C	GrN-14872						16480	150
16	C2	Con ¹⁴ C	GrN-14873						16850	150
17	B	Con ¹⁴ C	GrN-14874						16890	160
17	D1	Con ¹⁴ C	GrN-14875						17050	160
17	EA	AMS	UtC						19600	300
17	EB	Con ¹⁴ C	GrN-14876						16940	160
17	F	Con ¹⁴ C	GrN-14877						16790	160
18	B2	Gas counting	F-43	Charcoal	1972-1973	ABA	Azzi 1974	1971	14260	280
18	B2	Gas counting	F-44	Bone	1972-1973	HCl	Azzi 1974	1971	20200	305
18	B2	Con ¹⁴ C	R-1324						17200	150
18	B3	Gas counting	F-45	Charcoal	1972-1973	ABA	Azzi 1974	1971	20160	310
19	A	Gas counting	F-46	Charcoal	1972-1973	ABA	Azzi 1974	1971	20730	290
20	B	Gas counting	F-47	Charcoal	1972-1973	ABA	Azzi 1974	1971	21260	340
20	C	Gas counting	F-48	Charcoal	1972-1973	ABA	Azzi 1974	1971	22220	360
20	C	Gas counting	F-49	Bone	1972-1973	ABA	Azzi 1974	1971	22110	330
20	DE	Gas counting	F-50	Charcoal	1972-1973	ABA	Azzi 1974	1971	22630	390
21	A	Gas counting	F-51	Charcoal	1972-1973	ABA	Azzi 1974	1971	23040	380
21	B	Gas counting	F-52	Charcoal	1972-1973	ABA	Azzi 1974	1971	23470	370
21	C	Gas counting	F-53	Bone	1972-1973	ABA	Azzi 1974	1971	24210	410
21	C	Gas counting	F-54	Charcoal	1972-1973	ABA	Azzi 1974	1971	23750	390
21	C	Gas counting	F-56	Bone	1972-1973	ABA	Azzi 1974	1971	24219	410
21	D	Gas counting	F-55	Charcoal	1972-1973	ABA	Azzi 1974	1971	24720	420
22	B	AMS	UtC-1412						26800	300
22	F4	AMS	UtC-1413						28300	800
23	A	AMS	UtC-1414						28100	400
23	B	AMS	UtC-1415						26300	400
24	A1	Con ¹⁴ C	UtC-1789	Coprolite					29300	500
24	B1	Con ¹⁴ C	UtC-1790						34300	900

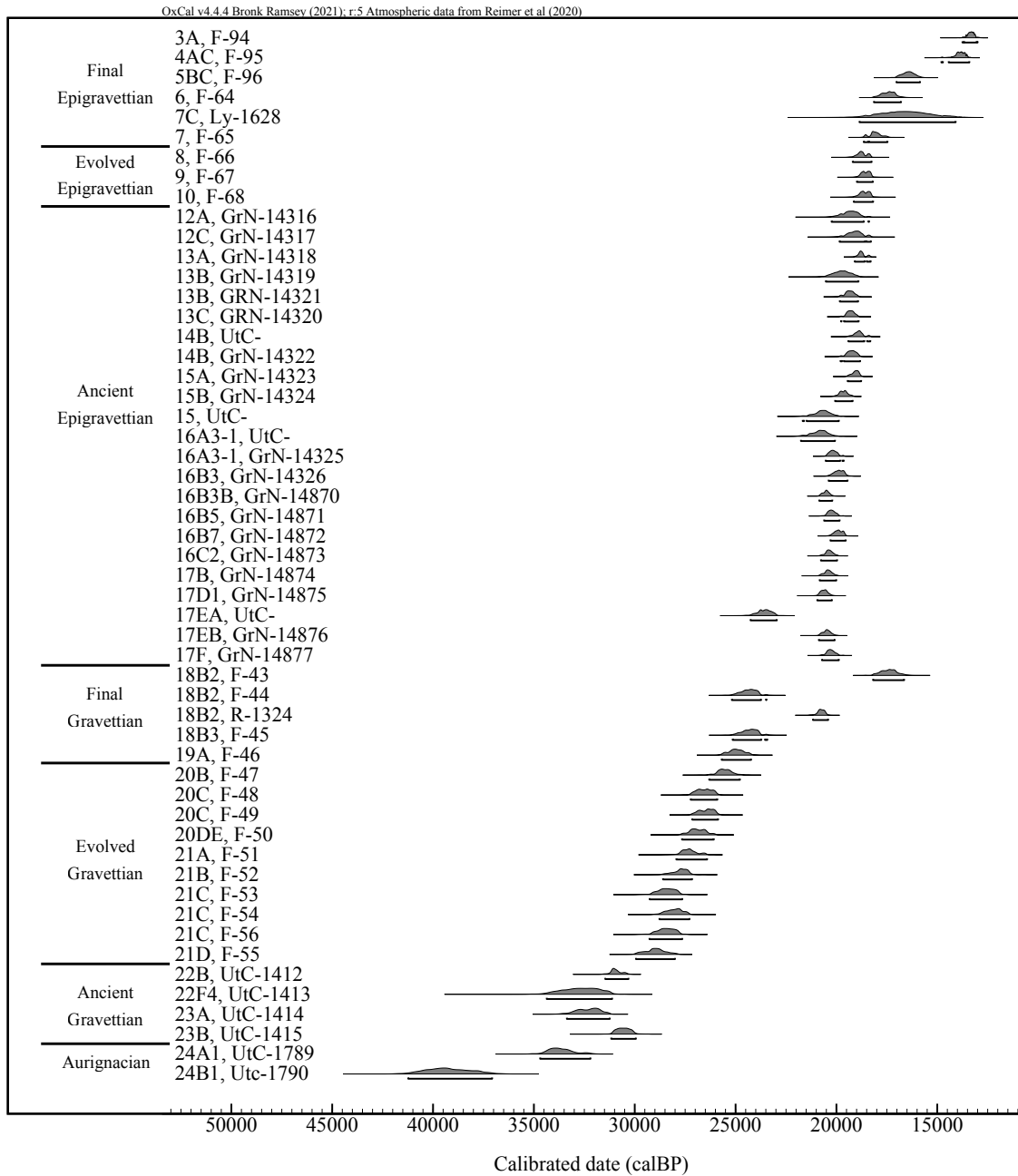


Figure 8.3: Previous radiocarbon determinations for Grotta Paglicci individually calibrated using IntCal20 in OxCal. Dates are labelled by their site layer and sublayer (e.g. 3A), as well as their lab code and number where known (F = Florence; Ly = Lyon; GrN = Groningen; UtC = Utrecht; R = Rome).

F-56, but according to [Azzi et al. \(1974\)](#), F-56 relates to a Roman house beneath Florence Cathedral, dated to 1660 ± 80 BP. The Paglicci date incorrectly listed as F-56 (24219 ± 410 BP) is probably a duplication of F-53 (Layer 21C, 24210 ± 410 BP) with a typo in the radiocarbon measurement. This error has been propagated through several publications ([Berto et al., 2017](#); [Boschin et al., 2018](#); [Ronchitelli et al., 2015](#); [Vermeersch, 2020](#)). Given that no documentation could be sourced for Groningen, Utrecht, or Rome laboratories, it was not possible to check for any similar inaccuracies in these measurements.

Inspection of the calibrated dates from Paglicci in stratigraphic order reveals inconsistencies and large uncertainties in existing radiocarbon measurements (Figure 8.3). Dates from Layer 18B2 in particular show dramatic differences to each other (F-43, F-44, and R-1324), likely due to varying degrees of contaminant removal or introduction during sample preparation and chemical pretreatment. Sample F-43, described by [Azzi et al. \(1974\)](#) as “appear[ing] contaminated” (11), has a ^{14}C age at least 2,000 years younger than expected, assuming a consistent and intact stratigraphy. F-44 is the oldest date for Layer 18B2, and therefore tentatively might be considered most reliable (see Chapter 5, Chapter 6, and [R. Wood, 2015](#)). UtC-1415 from Layer 23B also appears inconsistent with dates from adjacent Layers 23A and 24A1. Many dates for Layers 21, 22, 23, and 24 show large measurement errors (> 400 ^{14}C years) that could be improved using modern instruments, or at least constrained using Bayesian methods ([Ramsey, 2009a](#)). Layers below 18B2 also appear somewhat inconsistent with the site’s stratigraphy (e.g., 17F, 17EB, 17EA), but are not considered in this study.

Available documentation from the Florence laboratory suggests that charcoal and bone samples were all treated with a method somewhat similar to ABA (“Acid-Base-Acid”; Chapter 5), with ammonium hydroxide used as a base: “All samples undergo standard boiling pretreatment with 10% HCl, the sample is then leached with 3-5% NH_4OH , followed by another HCl treatment, and is finally washed with distilled water” ([Azzi et al., 1973: 479](#)). Again, no documentation on pretreatment methods was available for Groningen, Utrecht, or Rome laboratories.

A thin layer of volcanic material was found at the base of Layer 23C2, overlaying the Aurignacian deposits and red clay of Layer 24, and overlain by sandy gravels in 23C. No previous dates have been made on material from 23C, whilst adjacent Layers 23B and 24A1 show some chronological inconsistency, as described above and shown in Figure 8.3. Giaccio et al. (2006) correlate this tephra to the Plinian Codola eruption, and infer the age of the tephra as ~ 33 ky calBP based on the existing chronology presented in Figure 8.3.

Elsewhere, the Codola eruption has been correlated to TM-16b at Monticchio, which would place the eruption at ~ 26.8 kyr BP and disagree with Giaccio et al. (2006)'s estimate (Wulf et al., 2004). Stratigraphic and geochemical questions relating to Wulf et al. (2004)'s conclusion are highlighted by Giaccio et al. (2006), leading the authors to instead correlate the 23C2 tephra at Paglicci with the Tyrrenian and Adriatic marine tephra, C-10. C-10 has a single related radiocarbon date on foraminifera from the marine core KET80-03, of 29300 ± 700 BP ($35212 - 31934$ calBP, 2σ ; Giaccio et al., 2008; Paterne et al., 1999). Giaccio et al. (2008) confirm that the C-10 correlates to the Codola eruption, overlaying the Y-3 tephra that elsewhere has been dated to 29300 ± 700 calBP, 2σ (Albert et al., 2019). Therefore, the C-10 appears closer to the likely age of layer 23C2 at Paglicci, but given the inconsistencies and large errors shown in Table 8.1 and Figure 8.3, this interpretation might be viewed with caution.

8.2.3 Palaeoenvironmental Context

Several relevant palaeoenvironmental records available for consideration alongside the archaeological record at Paglicci. As well as the NGRIP record, three high-resolution proxies of palaeoenvironmental change for MIS 3 and MIS 2 are considered, recovered at locations relatively close to Paglicci. All three are pollen records, taken at Monticchio (Allen et al., 1999), Ioannina (Tzedakis et al., 2002), and Tenaghi Philippon (Müller et al., 2011). These records were also used as validation for downscaling experiments in Chapter 2, and proxy alignment in Chapter 3. Their existing chronologies are discussed in Chapter 3. Pollen records from Lagaccione

and Valle di Castiglione also cover MIS 3 and have similar characteristics (Fletcher et al., 2010), but their sample resolution is lower (> 400 years per sample) and are not considered here.

Figure 8.4 compares the three pollen records with the $\delta^{18}\text{O}$ measurements from the Greenland ice core record between 60 and 20 kyr BP. The pollen records are plotted on the IntCal time scale, using the most up to date chronologies described in Chapter 3. For Monticchio, this is the varved chronology with tephra layers (Harrison et al., 2010). For Ioannina, the chronology is tuned using tie points in the pollen signals (Harrison et al., 2010). For Tenaghi Philippon, 60 radiocarbon dates, tephra layers, and various tie points are used (Staff et al., 2019; Wulf et al., 2018).

As discussed in Chapter 2 and Chapter 3, Figure 8.4 shows that much of the variability of MIS 3 and 2 seen in the Greenland record is not reflected in pollen records from southern Europe. Where similar variability is apparent, it is rarely aligned with that seen in Greenland. Abrupt warming during the transition between GS 15 and GI 14 at ~ 54 kyr BP is shown at Tenaghi Philippon, but this peak was manually aligned to the Greenland record as one of the tie points for the chronology. GI 14 appears relatively warm at Ioannina, but the opposite is shown at Monticchio. All three records show relatively cold conditions between GS 5 and GS 2 ($\sim 30 - 20$ kyr BP), in line with the Greenland record, but variability is apparent at Monticchio. Assuming DO patterns had large-scale impacts over Europe, the asynchronicity in vegetation dynamics shown in Figure 8.4 could be because of any one of a number of factors, such as sampling resolution, chronological error, or localised and nuanced ecological responses to hemispheric climatic variability (see Chapter 2).

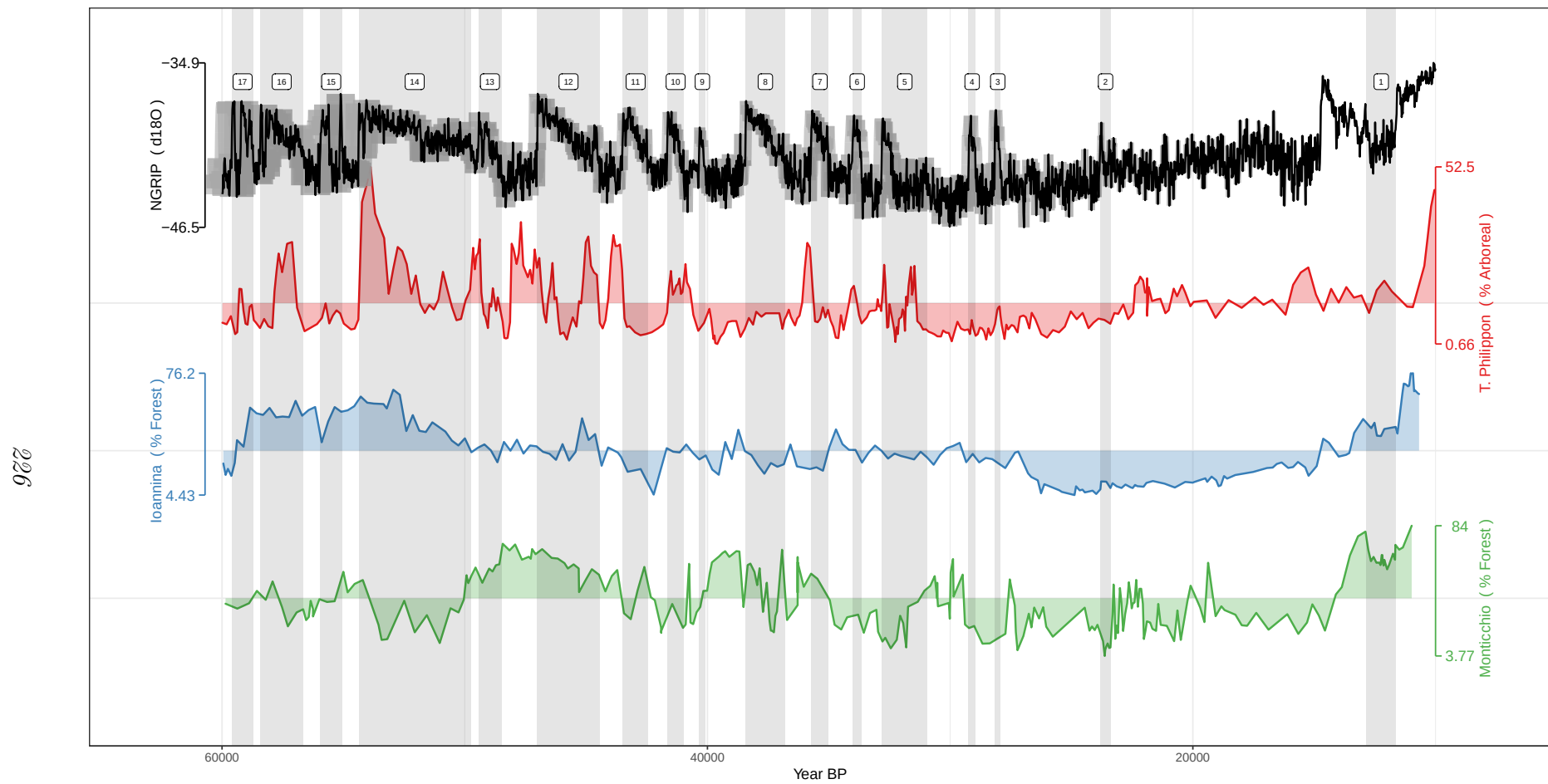


Figure 8.4: Environmental records for MIS 3. The NGRIP record (black line) and associated chronological uncertainty (grey shading) is shown above, converted to the IntCal time scale (Rasmussen et al., 2014). % arboreal pollen at Tenaghi Philippon (red; Müller et al., 2011), and % forest pollen at Ioannina (blue; Tzedakis et al., 2002) and Monticchio (green; Allen et al., 1999) are shown below. Interglacial periods defined by Rasmussen et al., (2014) are overlain with grey shading.

In parallel to these regional records, site-specific palaeoenvironmental reconstructions at Grotta Paglicci have been made by habitat suitability analyses of small mammal remains (Berto et al., 2017), proportional analysis of macromammal (Boscato, 2004; Boschín et al., 2018), and avifauna assemblages (Tagliacozzo & Gala, 2004).

Berto et al. (2017) examine bone fragments of small mammals from throughout the entire Upper Palaeolithic interior cave sequence (Layers 24 - 1). The authors infer a relatively stable cold continental climate for the Gargano Promontory over Layers 24 to 17, with dry open meadows and fragmented forest areas, suggested by the consistent presence of *Microtus arvalis* (common vole). In light of this apparent stability, Berto et al. (2017) highlight the environmental “peculiarity” of the south-eastern region of the Italian peninsula over the Late Pleistocene, suggesting the area may have served as a refugium for certain species during stadial periods (Berto et al., 2017: 38). However, low frequencies of small mammals in parts of the sequence, along with the large uncertainties attached to radiocarbon dates for Layers at Paglicci, mean that the stability inferred and comparisons between the small mammal record and records such as Monticchio or Greenland should be tentative.

Macromammal remains analysed by Boschín et al. (2018) similarly show fairly consistent abundances of *Bos primigenius*, *Equus ferus*, *Capra ibex*, *Rupicapra* sp., and *Equus hydruntinus* over Layers 24 to 18. High frequencies of *Capra ibex* and *Equus ferus* indicate open and dry environments throughout this period, consistent with the small mammal record. Variation in *Bos primigenius* and *Capra ibex* abundances are explained by changing hunting strategies, as supported by changes in lithic assemblages (Boschín et al., 2018; Wierer, 2013). Avifaunal studies are also largely in agreement with the small and macro mammal assemblages, analysed between layers 24 to 22 by Tagliacozzo & Gala (2004). High frequencies of *Pyrrhocorax graculus* (Alpine chough) and *Columba livia* (common pigeon) suggest steppe or grassland landscapes with areas of bare rock, whilst Layer 23B yielded boreal and alpine species indicative of colder conditions (Tagliacozzo & Gala, 2004).

Together, site-specific palaeoenvironmental reconstructions for Layers 24 to 18 suggest a colder, more arid and open landscape on the Gargano Promontory than today. According to [Boschin et al. \(2018\)](#) and [Berto et al. \(2017\)](#), these conditions appear to have been fairly stable through the Aurignacian and Gravettian, until around GS 2. Some variability is identifiable, possibly related to wider climatic oscillations, but these changes are by no means as dramatic as seen in the Greenland records, or indeed regional European records.

8.2.4 Summary

Regional palaeoenvironmental records are unclear in their relationship to archaeological changes at Paglicci, in part due to the uncertainties surrounding the site's existing dates, but also because of issues relating to their own respective chronologies (Chapter 3). Site-specific reconstructions at Paglicci seem to indicate relative stability, in contrast to many of the pollen signals in Figure 8.4. Asynchronous fluctuations in their signals may reflect their poorly constrained chronologies, or local nuances and specific ecological responses to large-scale climate variability. These factors are discussed in Chapter 2 and Chapter 3.

From the collection of radiocarbon dates from Paglicci and their documentation, it is clear that existing chronologies at the site have stratigraphic inconsistencies, large uncertainties, and are likely to benefit from Bayesian modelling to constrain their calibrated approximations and detect outliers. In short, the radiocarbon chronology at Paglicci needs updating. In the next section, methods for building a revised chronology at Paglicci are described, along with strategies for comparing the resulting chronology to relevant palaeoenvironmental records.

8.3 Materials and Methods

8.3.1 Radiocarbon Dating

8.3.1.1 Sample Selection

In 2013, 43 samples recovered from Layers 24 to 18 at Paglicci were collected from the University of Siena for sampling for radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). 33 of these samples were charcoal, and 10 were bone from Layers 23 to 20. From the 43 samples collected, 28 new radiocarbon determinations were made: 9 in 2013; and 17 between 2021 and 2023.

All new radiocarbon determinations presented in this study were made on charcoal samples. Initial %N analysis on the 10 bone samples from Paglicci suggested poor collagen preservation for all samples collected ($< 50\%$ N; [Brock et al., 2010](#)). Conditions for charcoal preservation, on the other hand, appear very favourable at Paglicci, with charcoal present in all Upper Palaeolithic layers. Visual inspection of all charcoal samples confirmed excellent preservation, with consistently large and clean fragments (i.e., single entities) of charcoal available for dating, often with minimal sediment adhesion.

Samples were selected for dating to provide a consistent frequency of dates across the stratigraphic sequence. Layers 18, 23, and 24 were prioritised due to stratigraphic inconsistencies and large uncertainties for previous radiocarbon measurements (Figure 8.3, Table 8.1). Focusing on the earlier part of the sequence (Layers 23 and 24) also allowed for a more detailed investigation of the earliest Aurignacian and Gravettian at Paglicci, and their relationship with environmental change.

8.3.1.2 Sample Pretreatment and AMS Measurement

Several pretreatment methods were used on charcoal samples. These were: Acid-Base-Acid (ABA); Acid-Base-Oxidation-Stepped Combustion (ABOx-SC); and Acid-Oxidation-Stepped Combustion (AOx-SC). For the 9 samples dated in 2013, either

ABA or ABOx-SC was used, with the latter applied to samples from earlier layers (below layer 22).

In 2013, ABOx-SC was the predominant pretreatment protocol used at the ORAU for charcoal samples suspected to be older than ~25 ky BP (Brock et al., 2010). Internal experiments showed that a procedure omitting the base step from ABOx-SC (AOx-SC) produced indistinguishable results to ABOx-SC, resulting in far fewer sample failures during pretreatment chemistry, as discussed in Chapter 6 (R. Wood & Brock, 2011). Since 2016, AOx-SC has been the routine charcoal pretreatment method for samples older than ~ 25 ky BP. As such, all samples dated between 2021 and 2023 followed the AOx-SC method, rather than ABOx-SC.

ABA, ABOx-SC, and AOx-SC all result in a chemically treated and freeze-dried product that is combusted by the routine protocol at the ORAU. Detailed description of the ABA, ABOx-SC, and AOx-SC protocols at the ORAU are given in Chapters 5 and 6.

8.3.2 Bayesian Modelling

The discrepancies shown between previous radiocarbon dates were assumed to be due to contamination (Figure 8.3). From previous site descriptions, the integrity of the stratigraphy at Paglicci was assumed to be reliable (see Section 8.2). Stratigraphic ordering described by Palma di Cesnola (2006) was employed to build two Bayesian models: a Full Model using previously published radiocarbon dates in addition to the new dates presented here; and a Partial Model using only new dates. The OxCal 4.4 platform was used to build both models (Ramsey, 2009a). Dates were calibrated and converted to calendar years (calBP) using the most recent iteration of the Northern Hemisphere terrestrial calibration curve, IntCal20 (Reimer et al., 2020).

Stratigraphic information and radiocarbon likelihoods were combined in OxCal in *sequences*, separated by *boundaries* and comprised of uncalibrated radiocarbon dates and *phases*. A *sequence* describe a series of ordered dates and/or phases, with their order used as prior information in Bayesian modelling. In this case,

the dates' order was given by site stratigraphy. In both models, *phases* were used for sections of the stratigraphy where ordering of dates was not known or predetermined (i.e., where multiple samples from a single layer were dated, or where a sample was dated multiple times). *Boundaries* were used to define and extract estimates for the beginning or end of sequences to approximate the timings for the emergences of archaeological industries (e.g., Early Gravettian). **Sequence**, **Phase**, and **Boundary** are all OxCal functions. The OxCal code for both models is provided in the Appendices.

8.3.2.1 Full Model

Existing and newly measured radiocarbon data were combined in a Full Bayesian model for Aurignacian and Gravettian Layers (24 - 18) at Paglicci. The Full Model was comprised of all dates from these layers listed in Table 8.1 and the 28 new dates presented here, along with their stratigraphic order. Boundaries were included between archaeological industries and their corresponding layers (e.g., “Aurignacian”) as defined by Palma di Cesnola (2006) (Figure 8.2).

8.3.2.2 Partial Model

The Partial Model used only new radiocarbon measurements from Paglicci, made between 2012 and 2023. Building two models allowed for: a) an assessment of the internal agreement of the new dates produced; and b) a comparison between models and the effect of including previous (and inconsistent) radiocarbon dates in the Full Model. As for the Full Model, boundaries were placed according to archaeological definitions for layers at Paglicci given by Palma di Cesnola (2006).

8.3.2.3 Outlier Detection

Outlier detection analysis is useful in Bayesian model building to objectively identify radiocarbon measurements that significantly disagree with other dates in a sequence and the overall model structure (Ramsey, 2009b). Outlier detection was performed for both models, implemented in OxCal using a *t*-type outlier model.

The t -type outlier model is recommended for most archaeological purposes ([Ramsey, 2009b](#)), implemented in OxCal with assumption that the distribution of outliers can be approximated by a long-tailed Student's t distribution. Outlier analysis is performed in OxCal by the command `Outlier_Model()`, with the parameters *distribution* and *scale* defining the distribution and scaling of outliers, respectively. Here, the parameter *distribution* was defined as $T(5)$, implying a Student's t distribution with 5 degrees of freedom, and *scale* was defined as $U(0, 4)$, implying a uniform distribution between 10^0 and 10^4 .

For both models, the prior probability of each date being an outlier was varied according to the pretreatment method used and how recently the radiocarbon measurement was made. Very high confidence was assigned to ABOx-SC and AOx-SC dates, set at 1% prior outlier probability. This is because both methods are considered the most robust pretreatment methods currently practiced at the ORAU for dating charcoal, as discussed in Chapter 5 and Chapter 6. ABA dates were set to 20% outlier probability, as the ABA method has previously been shown to be problematic for some Upper Palaeolithic contexts older than 25 ky BP ([Higham et al., 2011](#); [R. Wood et al., 2012](#)). For the Full Model, all previously published radiocarbon dates from Paglicci were set to 50% prior outlier probability, given their poor documentation and clear stratigraphic inconsistencies (Section 8.2.2, Table 8.1, Figure 8.3). This variable outlier strategy allowed for sufficient flexibility in both models and reflected the analytical concerns with many of the previously published radiocarbon dates, with the most confidence placed in ABOx-SC and AOx-SC determinations.

8.3.3 Dating the 23C2 Tephra

[Giaccio et al. \(2006\)](#)'s 33 ky calBP age estimate for the 23C2 tephra was not included in either Bayesian model. This is because the estimate is inferred by just two chronologies: the existing estimate for Layer 23C2 at Paglicci; and a single

radiocarbon date from the KET80-03 marine core. Including the former as chronological information in the updated model would be circular, as this estimate was based on the previous chronology that was hoped to be tested, and possibly revised.

As for the latter, although the C-10 radiocarbon date agrees with the previous estimate of 33 ka calBP for 23C2 when calibrated with an updated marine calibration curve ($\Delta R = -125 \pm 80$), it is most logical to produce an updated chronology for the Paglicci that is independent of the C-10 date, then compare the results to verify the C-10 correlation. Rather than use the tephra as additional chronological information, the site stratigraphy was used to extract an independent age estimate for the 23C2 tephra in OxCal, and compared to the C-10 date and [Giaccio et al. \(2008\)](#)’s estimate (see Section 8.3.2). In both the Full and Partial Bayesian models, the boundary “Early Gravettian Start” was used to extract the chronological estimate for the 23C2 tephra, being equivalent to the position of the tephra layer.

8.3.4 Comparison to Palaeoenvironmental Proxies

The chronostratigraphic frameworks provided by both models were compared to $\delta^{18}\text{O}$ measurements from Greenland ice cores (NGRIP; [Rasmussen et al., 2014](#)), along with pollen measurements at Monticchio, Ioannina, and Tenaghi Philippon. DO events and the timing of interglacial periods were also considered, taken from the [Rasmussen et al. \(2014\)](#) chronology. Estimates for the durations of archaeological industries were extracted from the Bayesian model, then compared to palaeoenvironmental signals in order to examine the relationship between the timing of archaeological changes and the broader environmental shifts occurring over Europe.

The relationship between archaeological and palaeoenvironmental records was compared qualitatively, simply by plotting extracted durations from both Partial and Full Bayesian models alongside the environmental records. Specifically, attempts to quantify or calculate any relationship or “correlation” between environmental and archaeological records were avoided, because:

1. As discussed in Section 8.2.3 and Chapters 3 and 2, the nature, nuances, and uncertainties of the various records considered, as well as their chronologies, vary considerably. Although it is possible to “standardise” their chronologies to the same time scale for the purpose of plotting the data, the uncertainties of the underlying chronologies and data are obscured by doing so.
2. Although there are some inconsistencies in the previously published radiocarbon chronology, it is clear that many of the major archaeological phases at Paglicci (e.g., the Aurignacian) span periods of considerable northern hemispheric environmental variability (Figures 8.1 and 8.3). Even with the aid of Bayesian models, for most Upper Palaeolithic contexts, obtaining precise estimates for when a particular archaeological industry emerged is unlikely to be possible at the resolution necessary to confidently address the question of environmental causality. Therefore, in most cases it is more appropriate to assess the *duration* of archaeological phases.

In other words, although it is unlikely that we will be able to ascertain whether a given industry emerged during a specific stadial or interstadial, we can assess whether the industry persists through one or several DO cycles, or is short-lived in relation to other archaeological phases. As such, rather than attempt to align (or, “correlate”) the emergence of a given industry to a particular environmental event or period, the environmental backdrop during which the industry persists is considered instead.

3. Attempting to “correlate” probability density functions representing the emergence or duration of various archaeological phases, such as those extracted from OxCal, with a univariate time series of environmental change does not make sense mathematically.

Table 8.2: New radiocarbon determinations for samples from Grotta Paglicci. Calibrated age ranges are given at 95.4% confidence intervals for unmodelled dates, the Partial Bayesian model using only the new radiocarbon measurements made in this study, and the Full Model using all available radiocarbon dates from Paglicci.

OxA	Layer	Tr.	Date	$\pm$	%C	$\delta^{13}\text{C}$	Unmodelled		Partial			Full		
							From	To	From	To	C	From	To	C
OxA-27739	18B2	ABA	21080	100	46.60	-21.64	25686	25187	25845	23832	105.0	26059	25152	102.4
OxA-43286	18B3	AOx-SC	21946	63	79.80	-21.71	26373	25962	26365	25965	101.9	26376	25963	100.4
OxA-P33370	19A	AOx-SC	22647	63	78.46	-27.96	27225	26476	27171	26432	63.6	27170	26444	68.4
OxA-27551	19B1	ABA	22530	150	49.90	-20.92	27185	26434	27654	26462	99.4	27478	26521	96.4
OxA-P33369	19B3	AOx-SC	23626	79	84.09	-27.94	27898	27661	27923	26491	75.1	27900	27653	100.3
OxA-27738	20C1	ABA	24480	150	49.90	-20.40	29117	28347	28986	26888	92.1	29074	28030	90.3
OxA-27737	21A1	ABA	23860	130	50.30	-21.17	28400	27730	29450	27737	50.6	29994	28138	28.6
OxA-P33365	21A1	AOx-SC	24861	77	79.01	-28.33	29223	28872	29243	27557	78.3	29225	28880	102.4
OxA-43244	21D	AOx-SC	25790	120	70.00	-19.85	30282	29911	30255	29896	103.6	30266	29899	102.2
OxA-43243	22A	AOx-SC	26100	130	80.60	-20.23	30765	30070	30863	30116	78.8	31855	30131	54.1
OxA-43242	22B	AOx-SC	27960	130	81.10	-21.26	32770	31530	31809	30970	26.9	31873	31249	41.9
OxA-42166	22C	AOx-SC	27630	140	82.80	-25.28	31819	31229	31810	31375	106.9	31857	31404	97.9
OxA-42165	22C2	AOx-SC	27700	140	81.30	-24.98	31895	31259	31882	31521	111.7	31897	31544	104.6
OxA-43268	22D	AOx-SC	28140	120	85.20	-24.38	32862	31734	31984	31679	82.2	31981	31686	84.4
OxA-28383	22E	ABA	27900	180	57.60	-21.23	32804	31305	32097	31739	139.9	32096	31740	136.3
OxA-42164	22E	AOx-SC	28060	140	81.00	-25.08	32854	31630	32105	31754	142.0	32104	31755	142.5
OxA-43241	22F2	AOx-SC	28090	140	72.20	-25.47	32860	31669	32238	31818	134.4	32288	31794	136.1
OxA-27736	23A	ABOx-SC	28190	170	79.50	-24.71	32960	31746	32699	31929	108.8	32635	31841	128.2
OxA-42163	23B	AOx-SC	27920	140	79.60	-22.28	32745	31444	32877	32003	11.6	32830	31856	33.9
OxA-27735	23C	ABOx-SC	28580	220	69.60	-21.36	33612	32027	33226	32147	113.5	33116	31887	91.4
OxA-42162	24A	AOx-SC	28940	160	80.50	-21.62	34017	32957	33742	32985	108.3	33814	32950	106.4
OxA-27734	24A1	ABOx-SC	28980	240	42.00	-20.94	34208	32334	34085	33228	112.9	34080	33230	112.2
OxA-P33354	24A2	AOx-SC	29460	150	79.32	-27.86	34381	33690	34394	33793	105.1	34375	33748	104.6
OxA-P33375	24A4	AOx-SC	32350	200	75.20	-26.71	37106	36227	37050	36230	102.8	37056	36236	102.7
OxA-27841	24B1	ABOx-SC	33030	300	61.10	-24.58	38972	36796	38640	36742	110.5	38455	36706	113.1
OxA-42167	24B1	AOx-SC	33060	250	80.70	-25.20	38865	36908	38570	36857	109.2	38416	36821	111.5

8.4 Results

New radiocarbon determinations from Upper Palaeolithic layers at Grotta Paglicci are presented in Table 8.2. 26 successful new radiocarbon measurements were made from Aurignacian, Early Gravettian, Evolved Gravettian, and Final Gravettian

levels, by ABA, ABOx-SC, and AOx-SC methods. Unmodelled and uncalibrated radiocarbon ages range from 21080 ± 100 to 33060 ± 250 BP, which calibrate to between approximately 25500 and 37600 calBP. Unmodelled age ranges at 95.4% probability are presented in Table 8.2, alongside modelled approximations at 95.4% probability from both Partial and Full Models. Unmodelled dates are also plotted in Figure 8.5.

The ABA date for Layer 21A1 (OxA-27737) was inconsistent with the stratigraphic sequence, being younger than the ABA date for the overlying layer, 20C1. Because of this inconsistency, this sample was redated by AOx-SC, producing a date in better agreement with the site's stratigraphic order (OxA-P33365), with marginally lower error. The AOx-SC date for Layer 22B (OxA-43242), meanwhile, was older than would be expected for its stratigraphic position, but with overlapping error with underlying layers. This might suggest insufficient contaminant removal by AOx-SC for samples from underlying layers 22C and 22C2. The AOx-SC date for Layer 23B (OxA-42163) was also inconsistent with the site stratigraphy, but again, its radiocarbon determination overlaps with errors of dates from adjacent layers. Other than these anomalies, calibrated and unmodelled radiocarbon dates were consistent with the site stratigraphy.

Table 8.1 also shows $\delta^{13}\text{C}$ and %C values for all samples, both routinely measured alongside ^{14}C determinations at the ORAU. All $\delta^{13}\text{C}$ values were typical of charcoal samples. Other than a low %C yield for OxA-27734, none of these measurements were unusual or fell outside of accepted ranges, which might have otherwise indicated problematic results. Overall, %C values for the 5 samples pretreated by ABA were somewhat low, but within acceptable ranges (Brock et al., 2010). High %C values for ABOx-SC and AOx-SC samples were typical for the method, with the exception being OxA-27734. The lower-than-average %C value for OxA-27734 was not explained by any laboratory observations or measurements, but perhaps was due adhering sediment that was not removed during chemical treatment. The radiocarbon determination for this sample, in comparison to results for samples from layers above and below it, was in line with what would

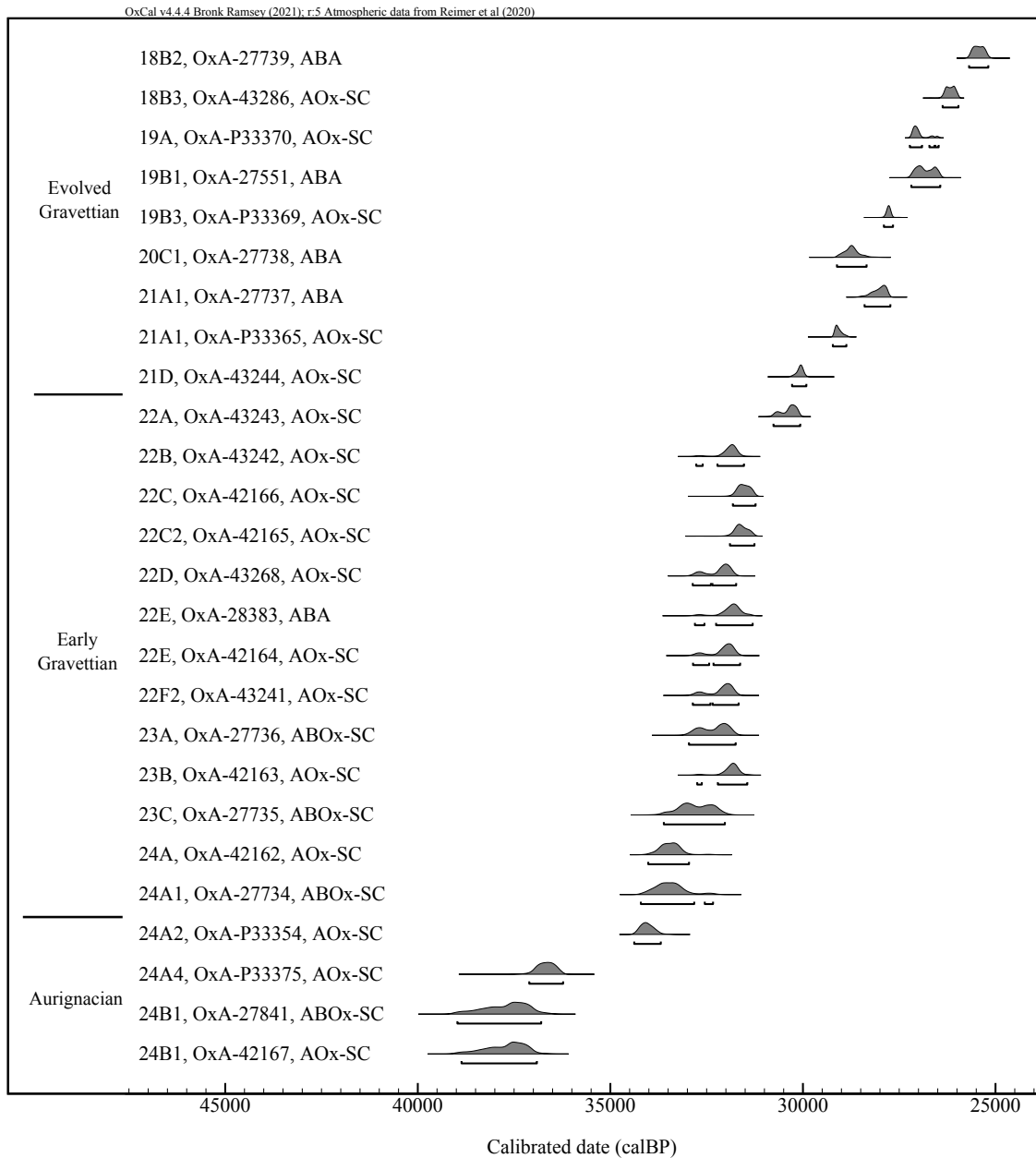


Figure 8.5: Unmodelled and calibrated radiocarbon measurements from Grotta Paglicci. Labels indicate sample layer (e.g. 18B2), OxA number, and pretreatment method (e.g. ABA). OxA-27737 from Layer 21A1 was suspected to be contaminated given its poor agreement with adjacent dates, therefore was redated using AOx-SC. The AOx-SC measurement (OxA-P33365) is more consistent with adjacent dates and the site stratigraphy.

be expected and does not suggest the sample was significantly contaminated by exogenous carbon.

%C yields for AOx-SC were much higher than for ABA and for ABOx-SC, as is typical for this method (see Chapters 5 and 6). All AOx-SC %C values were above 70%, compared to averages of 50.86% for ABA and 63.05% for ABOx-SC. These differences are likely to reflect differences in protocol, rather than variability in the carbon content of samples. Precombustion using O₂ for AOx-SC typically produces a much higher %C yield, as non-carbonic material is combusted and discarded in the precombustion process (again, see Chapters 5 and 6).

8.4.1 Full Model

The Full Bayesian Model that includes all new and previously published radiocarbon determinations is shown in Figure 8.6. Posterior probabilities for many previously published dates indicated poor agreement with new determinations by ABOx-SC and AOx-SC, and many outliers were detected by outlier analysis. The majority of outliers were previous determinations made at the Florence, Rome, and Utrecht laboratories. The only date produced for this study detected as an outlier in the Full Model was ABA date OxA-27737, which after repeating pretreatment by AOx-SC (OxA-P33365), showed good agreement with the model.

8.4.2 Partial Model

The Partial model, using only new radiocarbon determination, is shown in Figure 8.7. The only outlier in the Partial model is OxA-27737 (discussed above). Otherwise, the Partial Model showed excellent agreement with the site stratigraphy and ordering. Further comparison between Full and Partial Models is provided in Section 8.5.

8.4.3 Dating of the 23C2 Tephra

Posterior estimates for the age of the tephra layer found at the base of Layer 23CII, overlaying Aurignacian Layer 24A, range between 33454 and 32219 calBP

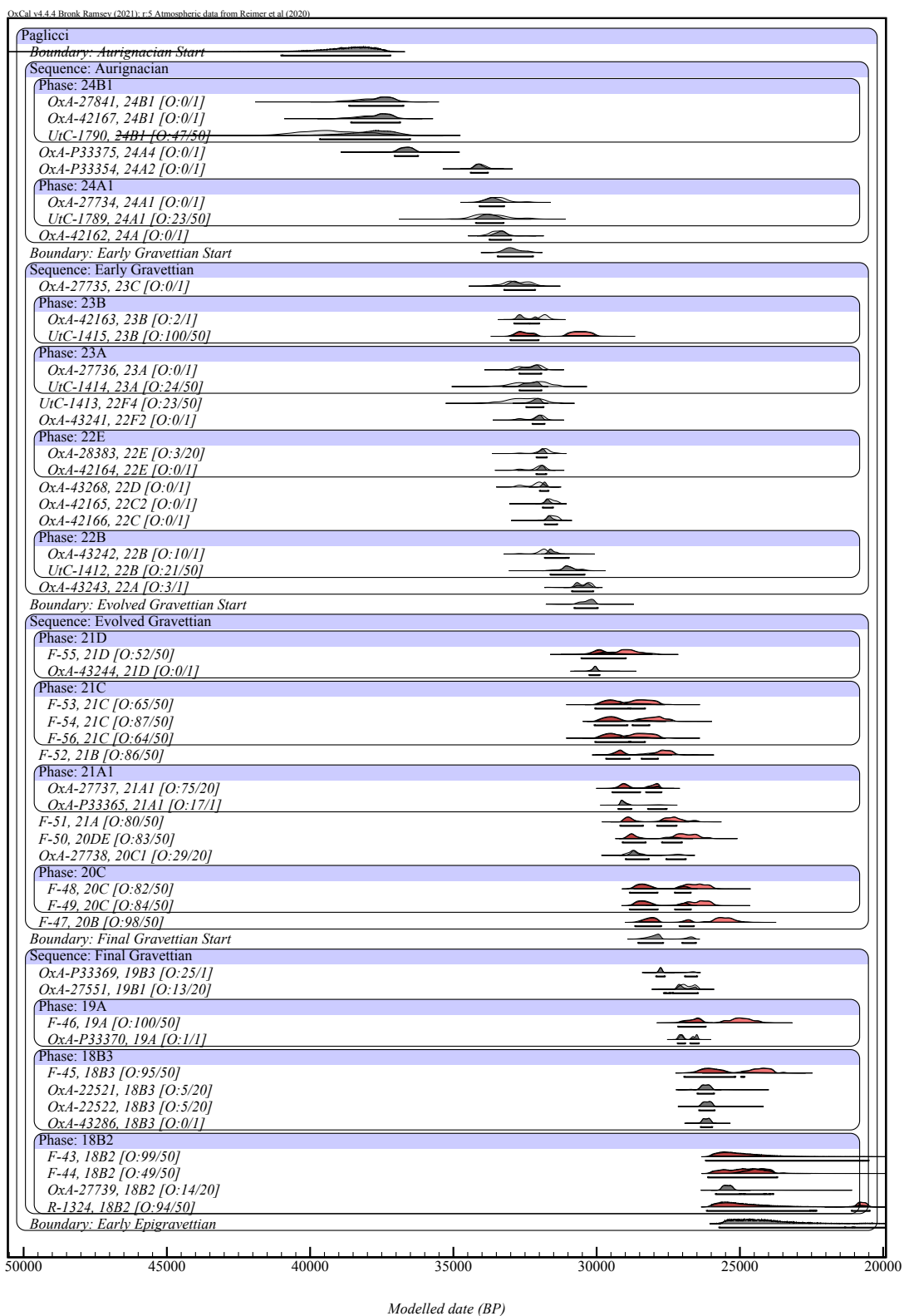


Figure 8.6: Full Bayesian model, built using new radiocarbon measurements and existing published dates from Grotta Paglicci. Detected outliers are highlighted in red.

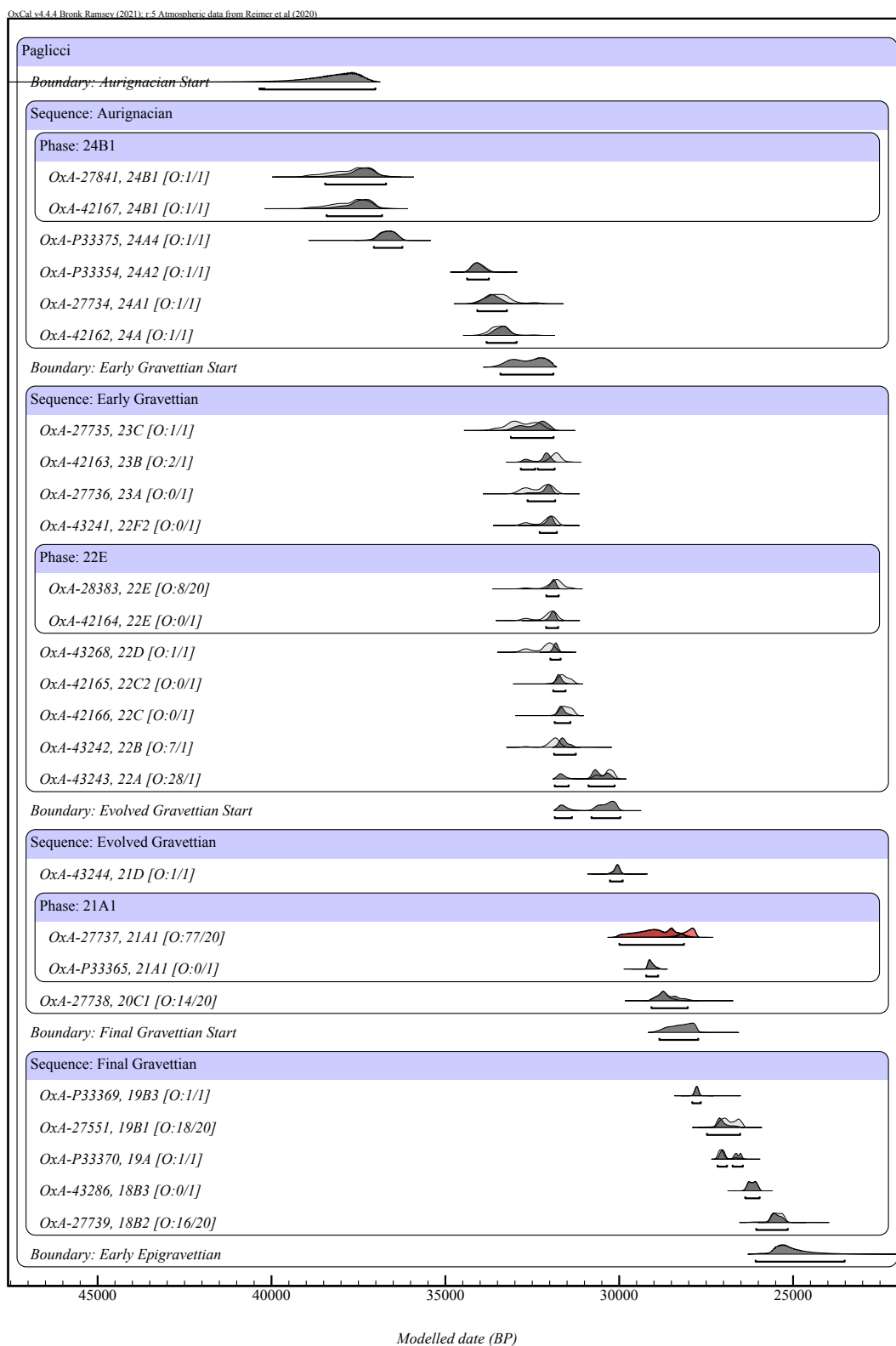


Figure 8.7: Partial Bayesian model, using only radiocarbon dates produced in this study. Detected outliers are highlighted in red.

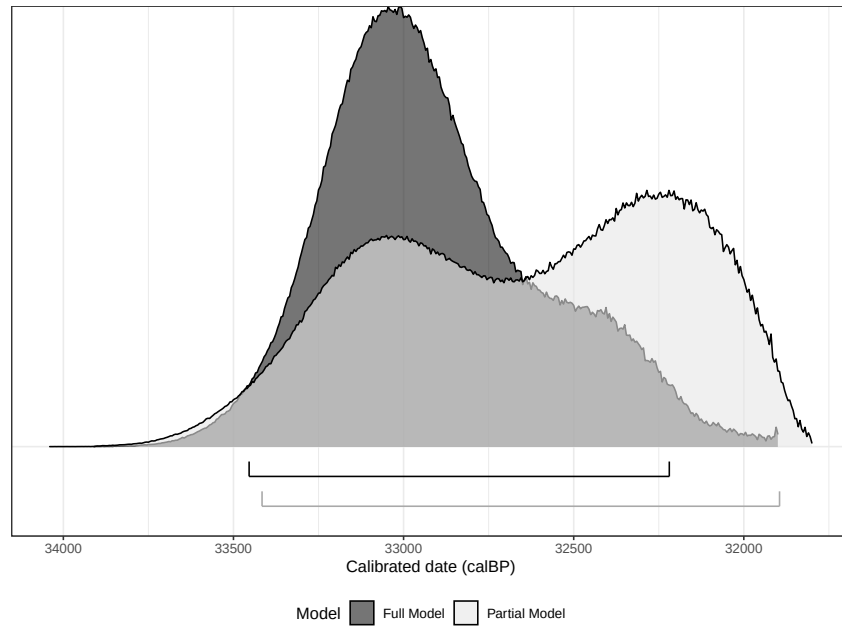


Figure 8.8: Age estimates for the 23C2 tephra layer, extracted from Full (dark grey) and Partial (light grey) models. 95.4% confidence intervals are also shown.

at 95.4% confidence intervals for the Full Model, and 33416 to 31895 for the Partial Model. Both are in agreement with [Giaccio et al. \(2008\)](#)'s correlation with the C-10 tephra, and with the foraminifera date from the KET80-30 marine core. These estimates are shown in Figure 8.8.

8.4.4 Comparison to Environmental Records

Figure 8.9 shows the durations of major archaeological phases at Paglicci, extracted from the Full and Partial Models. The NGRIP $\delta^{18}\text{O}$ record, GIs 8-2 and associated uncertainties, and pollen records from Tengaghi Philippon, Ioannina, and Monticchio are also shown. No clear or consistent relationship was identified between the emergence of archaeological industries at Paglicci and these records of environmental change. Estimates extracted from either model show that Aurignacian, Early Gravettian, Evolved Gravettian, and Final Gravettian archaeological industries all persisted during periods of pronounced regional climate variability, as implied by the NGRIP and pollen records shown.

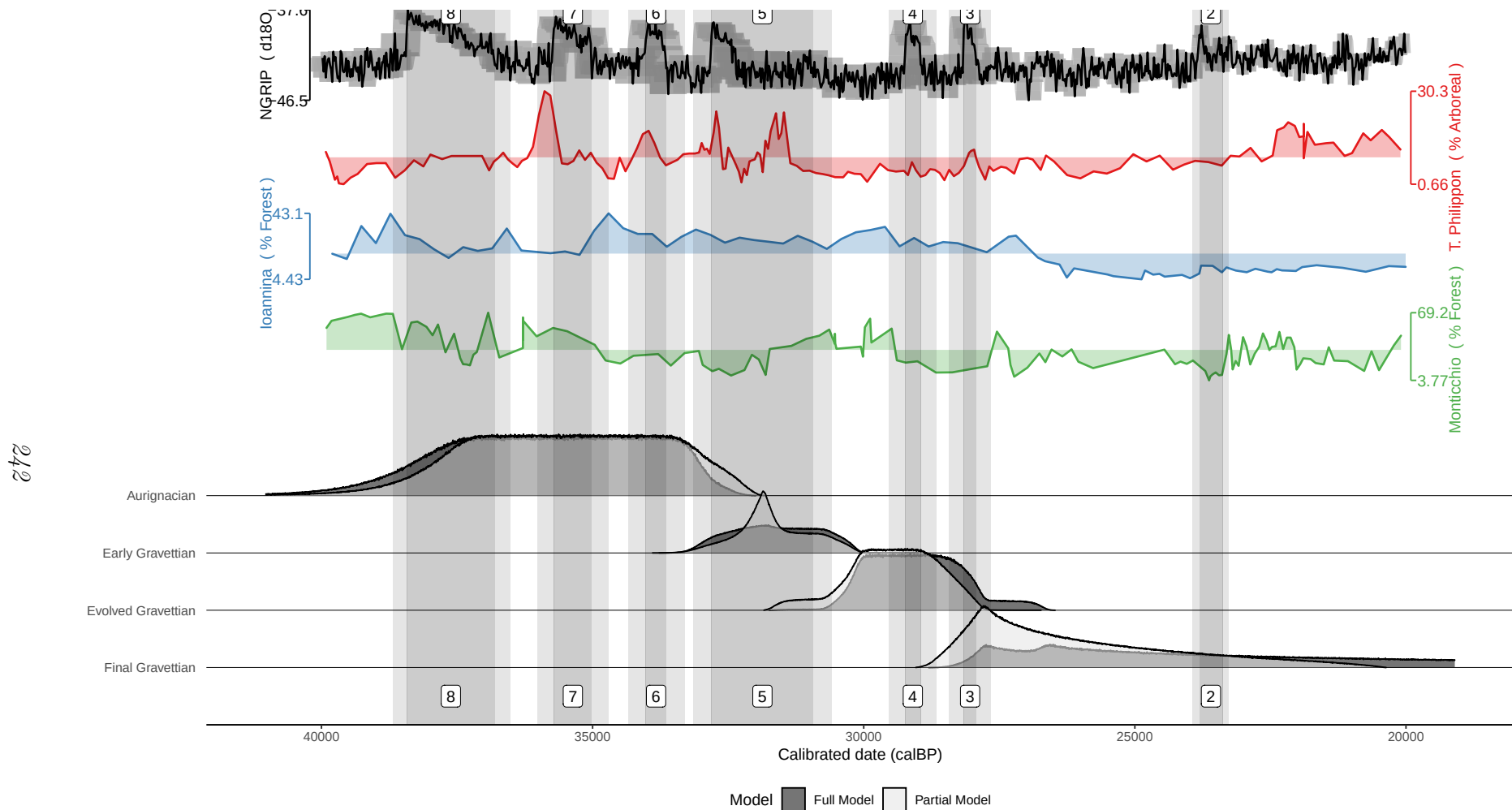


Figure 8.9: Estimates for the durations of major archaeological intervals at Grotta Paglicci, extracted from the Partial (light grey) and Full (dark grey) Bayesian models (bottom half). Above, NGRIP $\delta^{18}\text{O}$ values (black; Rasmussen et al., 2014), % Arboreal pollen at Tenaghi Philippon (red; Müller et al., 2011), % Forest pollen at Ioannina (blue; Tzedakis et al., 2002), and % Forest pollen at Lago Grande de Monticchio (green; Allen et al., 1999) are plotted for comparison. Greenland Interstadials 8 to 2 are shown in grey and labelled, along with their uncertainties (Rasmussen et al., 2014).

8.5 Discussion

This chapter presents an updated radiocarbon chronology for Aurignacian and Gravettian layers at Paglicci, using a series of new AMS dates on charcoal recovered at the site. The new radiocarbon dates strongly agree with stratigraphic position, decreasing in age from lower to upper layers, but many of the new dates show significant differences to previous radiocarbon dates from the same layers. Using the new determinations, two Bayesian models were built, with greater confidence given to ABOx-SC and AOx-SC dates by assigning them lower prior outlier probabilities in OxCal. In practice, this means that the chronologies produced by both models are broadly anchored around determinations produced by these methods.

The two models are largely in agreement, but some differences are apparent in their posterior estimates for modelled radiocarbon dates, and their estimates of durations of archaeological phases at the site (Figures 8.11 and 8.9). The results were qualitatively compared to the NGRIP $\delta^{18}\text{O}$ record and three relevant pollen records of environmental change. Comparisons to environmental records did not reveal any clear relationship between the emergence of various archaeological phases at Paglicci and environmental change. The Aurignacian, Early Gravettian, Evolved Gravettian, and Final Gravettian all persisted at Paglicci through stadial and interstadial periods, spanning multiple DO cycles (Figure 8.9).

8.5.1 Previous Chronologies at Paglicci

Figure 8.10 shows a comparison between previously published radiocarbon dates from the Florence, Rome, and Utrecht laboratories, and the new radiocarbon determinations made in this study, all from Layers 18 - 24 at Paglicci. The unmodelled calibrated dates suggest that many of the previously published dates underestimated the age of Upper Palaeolithic layers. Dates from the Florence laboratory for Layers 18 - 21 show a consistent offset compared to new dates, the latter being consistently older by between 1 and 2 kyr. Previous dates for Layer 18B2 are particularly problematic and vary significantly. Previous dates from lower

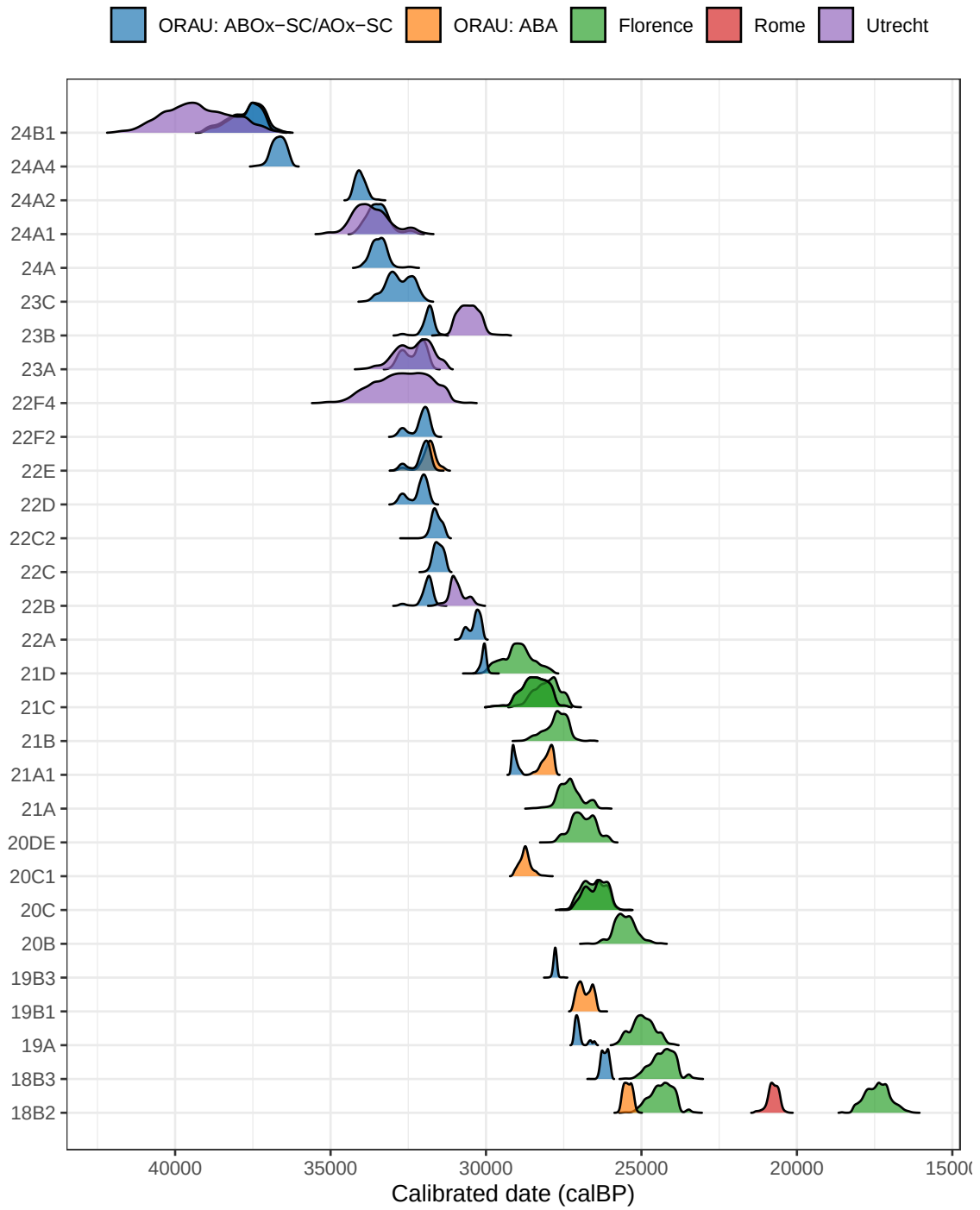


Figure 8.10: Comparison of existing and new calibrated dates from stratigraphic layers at Grotta Paglicci.

Layers (22 - 24), all from the Utrecht laboratory, show more agreement with the new radiocarbon determinations, but their uncertainties are much greater, with new dates generally showing much higher precision.

The likely cause of the significant differences between previously published dates

from Florence and Rome and the new dates produced in this study is insufficient contaminant removal during pretreatment for previous dates. Many of the Florence dates were pretreated using an ABA protocol, and as discussed in Chapter 5, in many Middle and Upper Palaeolithic contexts ABA has been shown to be less effective for contaminant removal in comparison to the more aggressive ABOx-SC and AOx-SC pretreatment methods. Although charcoal samples from Paglicci appear very well preserved and stable, comparisons between ABA (both newly produced or previously published) and ABOx-SC or AOx-SC dates for many layers show consistently older dates by ABOx-SC or AOx-SC (Figure 8.10). This suggests exogenous contamination may be present in charcoal throughout the sequence, and the ABA pretreatment does not remove this contamination as effectively as ABOx-SC or AOx-SC.

An alternative explanation for this pattern is that modern contamination was consistently introduced to ABA samples before measurement. For the Florence dates, this is possible. However, one sample from Layer 21A1 at Paglicci was dated twice at the ORAU, by both ABA (OxA-27732) and AOx-SC (OxA-P33365) methods. Samples are routinely dated at the ORAU alongside ^{14}C -dead “standard” samples, used to assess levels of background contamination. In both cases, the standards produced ages above 50 ky, indicating that no significant amount of modern contamination was introduced during either sample pretreatment. The AOx-SC result is around 1 ky older than the ABA date, suggesting improved exogenous contamination removal by more aggressive pretreatment protocols. Although this pattern is identifiable in several radiocarbon dates throughout the sequence (Figure 8.10), two samples from Layer 22E were also dated by both methods at the ORAU that produced very similar ages (27900 ± 180 BP for ABA; 28060 ± 140 for AOx-SC). This suggests the level of contamination at Paglicci is not consistent throughout the site’s layers, and that ABA might be sufficient for contaminant removal in some cases, such as for Layer 22E.

8.5.2 Bayesian Models

Bayesian statistical models are a strong analytical tool used in chronology building, allowing for incorporation of chronological information, such as radiocarbon determinations, with prior archaeological information, such as stratigraphic ordering. Improved constraints for chronologies can be produced as a result of Bayesian modelling, as well as the potential to obtain probabilistic posterior estimates for boundaries or events whose timings are unknown. Here, two Bayesian models were built and compared to assess the effect of including previous radiocarbon dates that were suspected to be unreliable. Modelled and calibrated estimates for all new radiocarbon dates for layers at Paglicci produced in this study are shown in Figure 8.11, for both Full and Partial Models.

Results confirm that the choice of using a variable outlier probability in both the Full and Partial Bayesian models is appropriate. Although the ABOx-SC and AOx-SC dates cannot be confirmed as the true age of the Paglicci charcoals, they are most likely to be the most accurate determinations of all available dates, given that they are consistently oldest of all methods (R. Wood, 2015). Because all the determinations from Rome and Florence appear to have underestimated the age of charcoals from Paglicci, as well as two from Utrecht (Figure 8.10), it could even be argued that 50% prior outlier probability is an optimistic prior estimate of the reliability of the previously published dates, and that their true outlier probability is higher.

The Full and Partial Models show agreement for most dates with very similar posterior estimates. Some layers show inconsistencies, such as for the lower part of the sequence where ABOx-SC and AOx-SC dates from Layers 23C, 23B, and 23A show differences in their distributions. Here, the Full Model shows wider estimates that span an older range than the Partial Model. This is probably due to the inclusion of UtC-1414 from Layer 23A, which has a large error that influenced the posterior distributions of dates from underlying layers (Figure 8.10). ABA date OxA-27737 from Layer 21A1 shows a wide distribution for Partial and Full

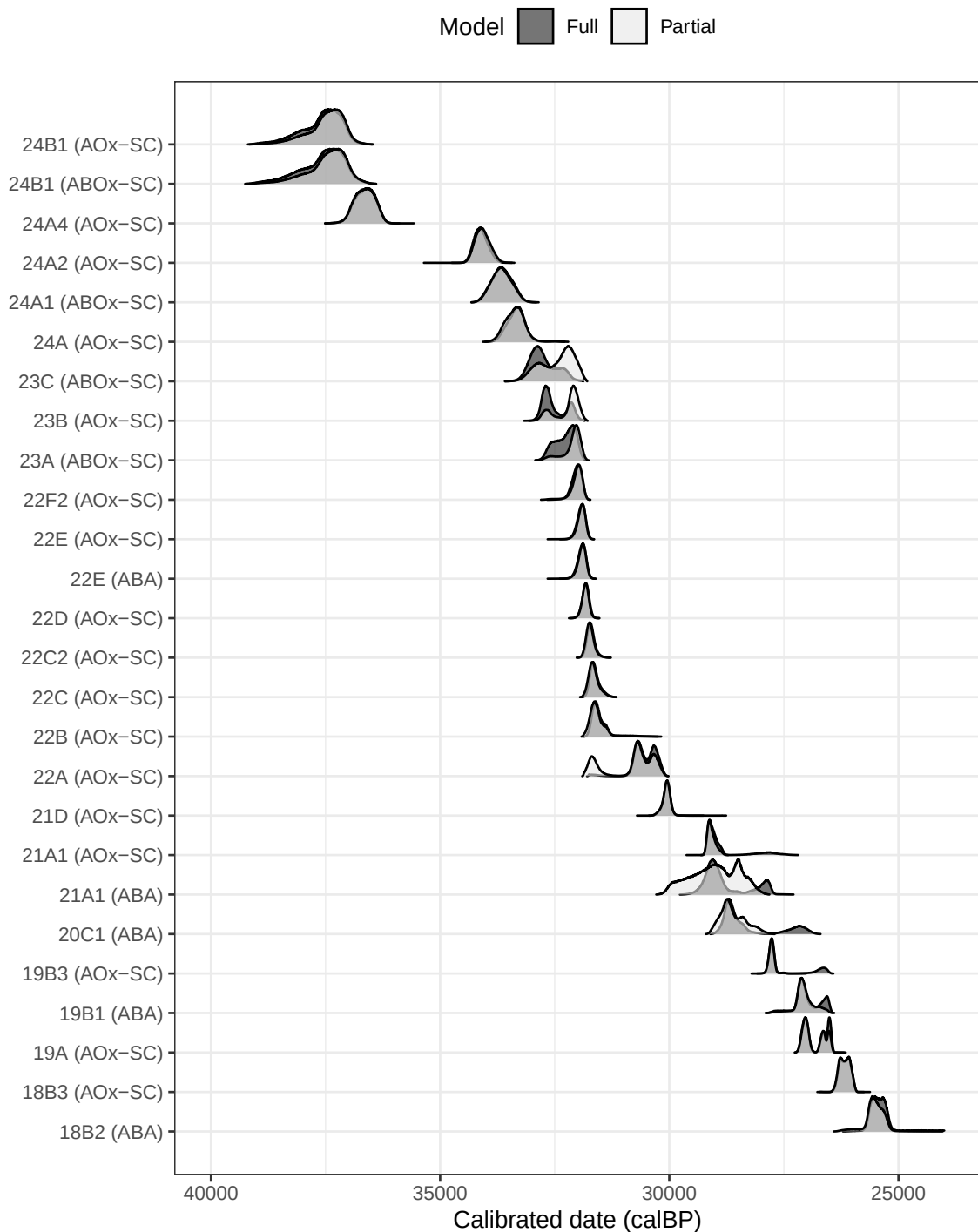


Figure 8.11: New radiocarbon determinations made in this study, calibrated and modelled by Full and Partial Bayesian models. Except some inconsistencies between probability densities across Layer 23 and 21A1, the two models show excellent agreement.

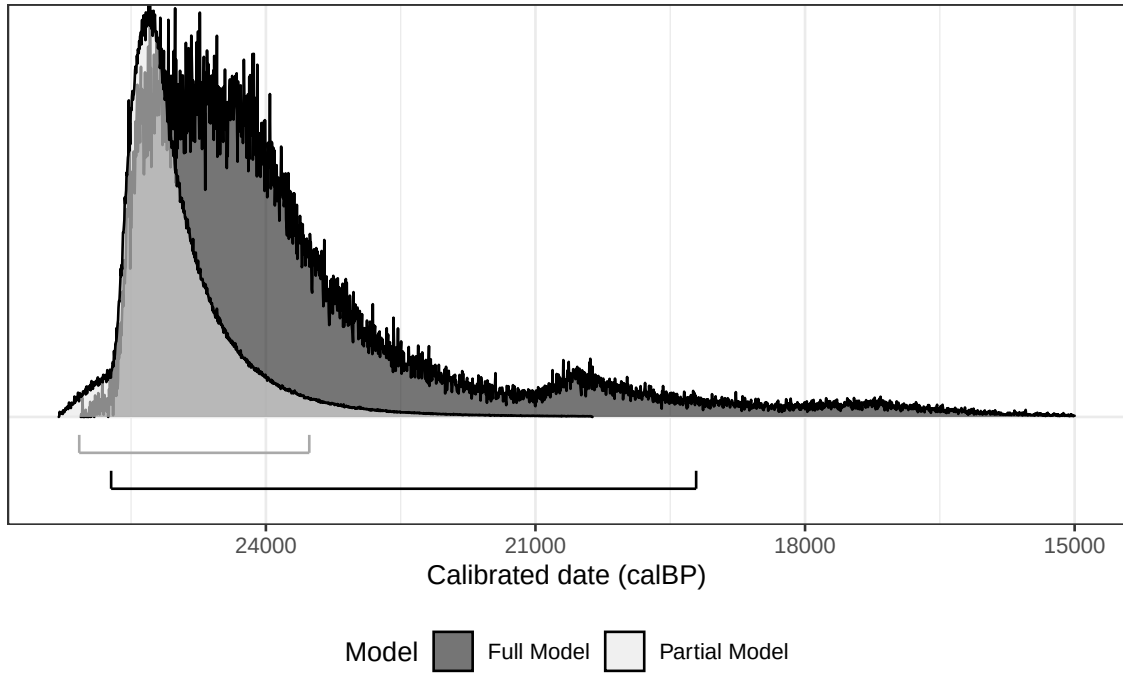


Figure 8.12: Estimates for the emergence of the Epigravettian at Paglicci from Partial (light grey) and Full (dark grey) models. Corresponding 95.4% confidence intervals are shown below for both models.

Models, but in both was highlighted as a potential outlier. The same sample redated by AOx-SC, meanwhile, shows excellent agreement between models (21A1 (AOx-SC) in Figure 8.11).

One of the greatest differences between Partial and Full Models is for estimates for the duration of the Final Gravettian (Figure 8.9), and corresponding emergence of the Epigravettian (Figure 8.12). Here, including previous dates with large uncertainties and poor agreement for Layer 18B2 (Phase 182B) in the Full Model created very long tailed estimates for the duration of the Final Gravettian, compared to the Partial Model. The boundary for the beginning of the Epigravettian (Epigravettian Start) similarly showed a poorly constrained estimate for the Full Model at 95.4% confidence interval (Figure 8.12; 25722-19212 calBP for the Full Model, 26077-23517 calBP for the Partial Model). No dates above Layer 18 were included in either model, given that the relative reliability of previously published dates could not be assessed by way of comparison to new dates for the same or adjacent layers.

Without more new dates for layers above Layer 18, the previous dates add

more uncertainty than value to estimates for upper boundaries of the sequence. Another reason for poor constraints on the upper boundary for the sequence is that OxA-27739 was assigned 20% prior outlier probability as an ABA date, a reduction here might constrain these estimates. Re-dating the sample from Layer 18B2 using ABOx-SC or AOx-SC methods could also be attempted, which might constrain the chronology while keeping the variability of outlier probabilities in both models consistent. Estimates for the start of the Aurignacian are similar for each model (Figure 8.9), where posterior distributions are “anchored” around ABOx-SC date OxA-27841 and AOx-SC date OxA-42167, both assigned very low prior outlier probability (1%).

The Partial Model (Figure 8.7) shows that the newly produced dates had excellent internal agreement. Given the number of outliers in the Full Model, and improved constraints for Final Gravettian duration and boundary in the Partial Model, the Partial Model is favoured in further discussion.

8.5.3 Estimates for the Emergence of the Aurignacian and Gravettian at Paglicci

The relationship between archaeological industries and the nature of “transitions” from one technocomplex to another are a central focus of many discussions in the archaeology of the Upper Palaeolithic. The chronological models presented here inform the archaeology of the “origins” of both the Aurignacian and Gravettian on the Italian peninsula, as well as the relationship between the two industries. Paglicci is unique among sites in Europe with archaeological sequences covering this period, in that its transition from Aurignacian to Gravettian is not separated by any apparent hiatus. Obtaining updated and precise estimates for the timing of the transition at Paglicci is made possible by new dates and the Bayesian models presented in this chapter (Figures 8.7 and 8.6). In turn, these models allows comparison to other sites with both Aurignacian and Gravettian industries (e.g., Abri Pataud, Roc de Combe, Willendorf II).

According to the Partial model, the Aurignacian at Paglicci persists from 38.5 to 32.3 ka calBP (95.4%), with the earliest age estimate for Layer 24B1 being between 38455 - 36706 calBP (Phase 24B1). A chronological gap is evident in the model between Layers 24A4 and 24A2 (Figure 8.7), but material has not yet been dated from the intermediate Layer 24A3. Dates for 24A3, if produced in the future, should be expected to lie between those of 24A4 and 24A2.

Estimates for the earliest Aurignacian at Paglicci are considerably later than Proto-Aurignacian levels at other Italian sites, such as Riparo Mochi, Fumane, Castelvita, and Serino. At Fumane, the earliest Proto-Aurignacian is dated to around 40.5 - 41.2 ka calBP (Higham, Brock, et al., 2009; Higham et al., 2011). Proto-Aurignacian layers at Castelvita and Serino are sealed by the CI tephra, which in combination with evidence from Fumane, implies that makers of the Proto-Aurignacian were present on the Italian peninsula by at least c. 41 ka calBP (Douka et al., 2014). Evidence from Riparo Mochi extends this further, where the start of the Proto-Aurignacian is placed at around 42.7 - 41.6 ka calBP (Douka et al., 2012). The Aurignacian “proper” appeared in southern Italy around 40 - 39 ka calBP at Serino, Castelvita, and Cala, around one thousand calendar years earlier than the earliest Aurignacian dates obtained from Paglicci.

These comparisons agree with Palma di Cesnola (2006)’s suggestion that the Aurignacian was produced by a separate population to the makers of the earlier Uluzzian, who did not arrive on the Gargano Promontory until after the Aurignacian appeared at sites further south, such as Castelvita and Serino. However, Layer 24B2 lies below all dated material at Paglicci. This layer yielded a small number of Aurignacian tools (Palma di Cesnola, 2004), which could suggest the earliest Aurignacian at Paglicci is in fact of similar age to that of other Italian sites. Dates from Layer 24B2 would confirm this, but unfortunately, no material was available to pursue this question.

The Aurignacian at Paglicci is directly succeeded by the Early Gravettian (Layers 22 - 23), lasting around 2700 calendar years from 32.9 to 30.2 ka calBP (95.4%). No transitional industry from Aurignacian to Gravettian levels is apparent at

Paglicci, but a gradual intensification of archaeological material through the Early Gravettian sequence is noted (Palma di Cesnola, 2006). The earliest Gravettian at Paglicci is later than at Fumane by around 2000 years (Higham, Brock, et al., 2009), but comparable to the dates for the first appearance of the Gravettian at Riparo Mochi (Douka et al., 2012), Rio Secco (Peresani et al., 2014; Talamo et al., 2014), La Cala (Boscato, 2007), and Fonte delle Mattinate. Douka et al. (2020) show that at Abri Pataud, a hiatus of between 1580 and 2940 calendar years separates the latest Aurignacian and earliest Gravettian material. At Paglicci, estimates for the earliest Gravettian Layer 23C and the latest Aurignacian from 24A overlap (Table 8.2; Figure 8.7), with no apparent hiatus.

8.5.4 Relationship Between Environmental Variability and Archaeological Changes at Paglicci

Comparisons of chronological estimates for the duration of major archaeological periods at Paglicci (Aurignacian, Early Gravettian, Evolved Gravettian, and Final Gravettian; Figure 8.4) to four relevant records of environmental variability (NGRIP $\delta^{18}\text{O}$; Tenaghi Philippon, Ioannina, and Monticchio pollen records) reveal that:

1. No clear relationship can be identified between the emergence (beginning) or persistence of any of the archaeological phases at Paglicci and large-scale environmental change during MIS 3 and 2.
2. The records of environmental change at Tenaghi Philippon, Ioannina, and Monticchio show no clear relationship to each other over MIS 2 and 3, based on their current chronologies.
3. Considering all records shown in Figure 8.9, the Aurignacian and Gravettian both persist through periods of pronounced environmental variability at Paglicci.

The Aurignacian and its arrival in Europe has frequently been associated with environmental factors, such as by explaining differences between its subdivisions as distinct eco-cultural niches (Banks et al., 2013), or discussing favourable environments and corridors for dispersal (Badino et al., 2020; Müller et al., 2011; Nigst et al., 2014). Results show that the Aurignacian at Paglicci, as for the Proto-Aurignacian at Fumane and Riparo Bombrini (Falcucci & Peresani, 2019; Riel-Salvatore & Negrino, 2018), cannot be attributed to one specific ecological “niche.” First, significant variability is apparent in the NGRIP record alongside the persistence of the Aurignacian at Paglicci; GIs and GSs 8, 7, and 6 all unfolded during this period. Clear fluctuations in arboreal taxa are also shown at Tenaghi Philippon, Ioannina, and Monticchio, although they bear no apparent relationship to each other. The Aurignacian at Paglicci, meanwhile, shows little variation, other than larger quantities of retouched bladelets in upper phases (Palma di Cesnola, 2006).

Similarly, the Gravettian at Paglicci cannot be attributed to one specific ecological period or set of conditions, despite claims for the Gravettian emerging as an adaptive response to cooling during MIS 3 (Taller & Conard, 2022). Here, the theory goes that resources are less evenly distributed in cooler climates, requiring a greater degree of mobility by hunter-gatherer populations. The portability of the Gravettian toolkit, its modular character, and its relatively fast spread have been forwarded as a response to this “subsistence pressure,” with greater mobility allowing for wider cultural contact and the exchange of ideas over larger areas (Taller & Conard, 2022: 102). These arguments are somewhat compelling, but evidence at Paglicci shows that the chronological resolutions of both archaeological and palaeoenvironmental records are not sufficient to attribute direct causality in this regard. The Early Gravettian at Paglicci, though persisting through GS 5, was also present during interstadial periods, and cannot be described as being “triggered and fuelled by the cooling climatic conditions from GS 7 onwards” (Taller & Conard, 2022: 102), nor as “innovative problem solving in connection with climatic stress” (Conard & Bolus, 2003: 363).

Discrepancies between regional and large-scale palaeoenvironmental records and site-specific palaeoenvironmental evidence at Paglicci cause further difficulties. At Monticchio, the closest record to Paglicci, abrupt fluctuations the % Temperate Forest signal are clear, but other European pollen records show completely different signals over this period. Reconstructions at Paglicci by [Boschin et al. \(2018\)](#) and [Berto et al. \(2017\)](#) suggest relative stability through MIS 3 and MIS 2 at Paglicci, although the resolutions of these analyses are relatively coarse.

At Bacho Kiro, [Pederzani et al. \(2021\)](#) discuss a similar discrepancy between local and regional proxies, where claims of environmental causality that the first arrivals of AMH in Europe was driven by a warming pulse and climatic amelioration (e.g., [Müller et al., 2011](#)) are undermined by stable isotope evidence. With a similar comment on the challenges of drawing parallels between distant environmental and archaeological records with poor resolution, the authors note that:

... direct evidence of cold climatic conditions presented here shows that an assignment of the IUP of Bacho Kiro to a GI is inappropriate. We believe that this apparent disagreement is caused by a combination of the large combined uncertainty of the chronometric dates of archaeological deposits and climatic records, spatial differences in the timing and expression of climatic events, and the different climatic and environmental features represented by each climatic proxy rather than methodological inconsistencies in any of the climatic reconstructions. Combined uncertainty of ice-core layer counting, ^{14}C , and U-series dates often span several thousand years, easily covering several climatic oscillations that can additionally vary in timing or intensity across the globe. ... The case of Bacho Kiro Cave again highlights that chronometric correlations between archaeological deposits and spatially distant climate records offer climatic context of a resolution that is not always appropriate for the questions asked. ([Pederzani et al., 2021](#): 5)

The questions of whether technological, behavioural, or cultural changes during the European Palaeolithic were direct “responses” to environmental variability cannot currently be answered by the data available. Instead, it is worthwhile to reframe these questions, toward a consideration of the *duration* of archaeological industries and patterns. In doing so, the variability seen in palaeoenvironmental

records is decoupled from its position as a “driver” of archaeological change. Instead, industries such as the Aurignacian or Gravettian are highlighted as resilient and adaptive, that persist through millennia of dramatic palaeoenvironmental variability (Figure 8.9).

From comparisons at Paglicci, it is clear that causal parallels between the archaeological record and regional palaeoenvironmental records are difficult to draw, primarily because of the nature of the evidence employed to do so. Aside from the issue of sampling resolution, before any conclusions of this kind are made the uncertainties associated with both archaeological and palaeoenvironmental data must also be acknowledged and considered. The Greenland ice core record is frequently referenced, plotted, and compared to archaeological evidence without any reference to its chronological uncertainty, for example (Badino et al., 2020; Marín-Arroyo et al., 2023; Taller & Conard, 2022). For MIS 3, this error is significant and severely limits the inferences that can be made when comparing specific archaeological horizons to the Greenland record, its stadials, and interstadials. The three pollen records considered here, meanwhile, have similarly significant chronological uncertainties, some based on clearly unreliable radiocarbon dates (e.g., Müller et al., 2011; Staff et al., 2019). A related problem surrounds the extent to which such chronologies can be assumed to be correct, or whether the approach of tuning their signals to each other and the Greenland record can be justified, by quantitative or qualitative methods (see Chapter 3). Resolving these chronological challenges is not straightforward, as the idiosyncrasies of each record and their chronologies records prevent “catch all” approaches to assimilating these data to a robust and coherent chronological framework.

The broad challenges relating to explaining archaeological shifts during the Upper Palaeolithic by regional environmental proxies are illuminated by this study. By comparing the new radiocarbon chronology at Paglicci to relevant ice core and pollen records, a number of problems become apparent that necessitate further research, data collection, or both, before confident relationships between the emergence of archaeological industries and environmental change can be drawn.

8.6 Conclusion

Radiocarbon chronologies at many Middle and Upper Palaeolithic sites have benefited from extensive dating campaigns over the last two decades, with significant progress made in the removal of contamination, improving the precision of dates, as well as their constraints by Bayesian methods. The new dates produced for this chapter for the Upper Palaeolithic sequence at Paglicci are further evidence that revised chronologies are necessary for many Palaeolithic sites, showing insufficient contaminant removal in previous dates, further highlighted by outlier detection in the Full Bayesian model.

Comparisons to palaeoenvironmental records show that ascribing causation to palaeoenvironmental deterioration or amelioration to explain the emergence of the Aurignacian and Gravettian on the Italian peninsula is not possible. This is not to say that these factors were not drivers of these changes. Rather, the available data are not sufficient in their resolution or precision to confidently draw parallels between their patterns. Instead of considering the emergence of archaeological industries alongside palaeoenvironmental records it is more appropriate to consider their duration.

Conclusion

This thesis has explored the palaeoenvironmental context of the archaeology of MIS 3 in Europe, exploring multiple numerical and laboratory-based questions in the fields of archaeological and palaeoenvironmental sciences. The broad themes of the research have been *Climate* and *Chronology*, contributing novel applications and new radiocarbon measurements to improve understanding of *how* dramatic and abrupt climate oscillations had impact over Europe, as well as *when* they occurred. Particular focus has been paid to DO events and their simulation and alignment in palaeoenvironmental records, as well as improvements to estimates for archaeological changes during the Palaeolithic. In doing so, the assumption of causality between large-scale environmental changes and behavioural, technological, and cultural change in Palaeolithic Europe has been problematised. Together, the chapters of this thesis have shown that uncertainties related to the *precision*, *accuracy*, and *resolution* of many records for MIS 3, be they archaeological or palaeoenvironmental, preclude an assessment of causality between these processes.

The aim of this thesis has been to improve chronological estimates for both archaeological and palaeoenvironmental changes during the Late Pleistocene. The former has been addressed by updated radiocarbon chronologies at Paglicci in Chapter 8. The latter has been shown to be a much more challenging endeavour. While Chapter 3 provided revised estimates for the timings of abrupt changes at Tenaghi Philippon, Monticchio, and Ioannina, they disagreed with existing age models for these palaeoenvironmental records. Meanwhile, Chapters 5 and 6 showed that radiocarbon chronologies can be susceptible to contamination, particularly at ages greater than 30 kya. It has been demonstrated that the design

and implementation of appropriate strategies for chronology building are context-specific; there are no “catch all” methods to alleviate these challenges across palaeoenvironmental records.

Each chapter of this thesis, with the exception of Chapter 7, has been presented as an independent study. In each chapter, the limitations of the respective study were considered and discussed. In Chapter 2, these related to validation records for MIS 3, and poor predictive performance for downscaling precipitation. Many more validation records could be considered and compared to downscaled predictions, while a more robust metric by which to quantify agreement between records could be employed. In Chapter 3, theoretical limitations relating to the assumption of synchronicity between records were considered, as well as more technical problems with the experiment design. Chapter 4 explored an alternative method for cross dating, but only one data set was used to test it. Experimental protocols designed in Chapters 5 and 6 were limited by small sample sizes and the absence of any known-age horizons at Fumane. The radiocarbon chronologies produced in Chapter 8 were limited by excluding Epigravettian dates, or not redating material from these layers, despite clear stratigraphic inconsistencies in previous radiocarbon determinations. Many of the studies presented are preliminary investigations that are limited in sample size and scope, intended as initial experiments in methods in alleviating some of the challenges of synthesising archaeological and palaeoenvironmental evidence. Together, these limitations reflect the scale of the challenge of building a coherent reconstruction of past climate variability in Europe, as well as assessing its effect on human populations.

In the introduction to this thesis, six objectives of this research were identified. Below, these objectives are reconsidered in light of the research that has been presented.

Objective 1

The first objective of this research has been to develop robust statistical methods to downscale GCM simulations and reconstruct the spatial impact of DO events

across Europe during MIS 3. In Chapter 2, a continuous GCM simulation was downscaled through a Machine Learning pipeline, with spatial cross-validation and automated feature selection methods. Cross-validation indicated severe spatial biases in coastal regions relating to the coarse resolution of the GCM. Alleviating this constraint in statistical downscaling requires either higher resolution GCM simulations, or limiting downscaling projections to areas without complex and coastal topography. Poor predictive performance for precipitation was also identified in this process, due to the complex and non-linear nature of this variable. In the absence of improved representation of precipitation patterns in GCMs, more complex topographic variables able to capture these complex relationships may improve predictive performance.

Objective 2

The second objective of this thesis has been to evaluate the predictions made by the downscaled GCM simulation by comparison to palaeoenvironmental records. In validating the downscaled temperature and precipitation predictions derived in Chapter 2, significant discrepancies in model-data comparisons were revealed. Here, the *precision* and *resolution* of downscaled GCM surfaces have been shown to be ample for archaeological questions, but the *accuracy* of these predictions could not be readily confirmed. Evaluating simulations of MIS 3 by way of comparison to palaeoenvironmental records requires that the latter's chronologies are well-constrained, are constructed with minimal uncertainty, and that the relevant signal has an adequate sampling *resolution* for comparison to modelled estimates. For MIS 3, the the validation records considered were not sufficient many of these requirements. As a result, the reliability of GCM-derived predictions and their application to archaeological problems has been questioned. Improved strategies for validation are essential before GCM-derived predictions can be understood with confidence.

Objective 3

The third objective has been to adapt and apply signal processing algorithms to assimilate modelled and observed data and produce new chronologies of palaeoenvironmental change over MIS 3. This was achieved in Chapter 3, where palaeoenvironmental observations from three key pollen records were synchronised to down-scaled predictions, by the signal processing algorithm DTW. In addition, a novel application of the MVM algorithm was explored and shown to have significant advantages for palaeoenvironmental synchronisation. The resulting realigned age models showed differences to their existing chronologies, calling into question either the reliability of existing chronological estimates, or the assumption that ecological shifts in Europe represented by these records were synchronous with temperature changes over Greenland during MIS 3. In alignment by DTW, the *accuracy* of existing age models for palaeoenvironmental proxies has again been questioned, as well as challenges relating to their *resolution*. For the archaeology of MIS 3, these uncertainties must be acknowledged before these palaeoenvironmental records are used to ascribe causality with any confidence.

The application of DTW to an altogether separate problem, of cross dating in dendrochronology, was explored in Chapter 4. The algorithm was shown to have clear potential in this field, where, unlike pollen records over MIS 3, the *resolution* of tree ring sequences is exceptionally high. Alignment by DTW outperformed alignment by Pearson correlation, the latter being the dominant method in dendrochronology. Comparisons showed that DTW's performance was particularly good for cases where missing or added rings were present in samples. The *accuracy* of the method was confirmed in this experiment, but further developments to the algorithm are necessary before the *precision* of DTW in dendrochronology can be improved. This chapter clearly shows that DTW could be an excellent alternative to Pearson correlation for cross dating, providing the preliminary foundation for future development of the algorithm for deployment in dendrochronology.

Objective 4

The fourth objective of this research has been to explore and design new methods for radiocarbon pretreatment. Chapters 5, 6, and 7 presented and discussed experiments in radiocarbon dating, where novel pretreatment methods were applied to various charcoal samples produced at the Palaeolithic site of Grotta di Fumane during MIS 3. In Chapter 5, a comprehensive comparison between ABA, ABOx-SC, and ABA-SC was provided, giving a quantitative estimate for the effect of precombustion in pretreatment. This chapter represents the largest set of comparative dates for ABA, ABOx-SC, and ABA-SC pretreatment methods. Chapter 6 expanded upon these findings, exploring a novel pretreatment method using an alternative and less-aggressive oxidation reagent under ABOx-SC. Neither method was able to remove contamination as effectively as ABOx-SC, but both experiments provide further illumination on *how* ABOx-SC is so effective. In the wider context of this thesis, both experiments have shown that the *accuracy* of a radiocarbon date is not guaranteed, particularly for the archaeology of MIS 3, with considerable variability across measurements shown.

Objective 5

The fifth objective of this thesis has been to produce new radiocarbon dates by the most robust pretreatment methods to provide updated chronologies of Palaeolithic cultural change in Europe. Chapter 8 presented a new series of AMS dates on charcoal from the Upper Palaeolithic site of Grotta Paglicci, Italy provided an updated chronology of Palaeolithic cultural change through MIS 3. Many previous radiocarbon estimates at Paglicci were shown to have suffered from insufficient contaminant removal by outlier analysis, and the new estimates were further constrained by Bayesian modelling. The updated chronology represents one of the most up-to-date and robust chronologies of cultural shifts during the Upper Palaeolithic in Europe.

Objective 6

The sixth and final objective of this thesis was to compare estimates for the timings of Palaeolithic cultural change to palaeoenvironmental records, and assess the relationship between these processes. The updated chronology provided an opportunity to explore this relationship, between cultural transitions at Paglicci and the broader climatic oscillations in Europe, the latter evidenced by several palaeoenvironmental records also considered in Chapters 2 and 3. Through the comparison made at Paglicci, the recurring conclusion of this research was confirmed: the *precision*, *accuracy*, and *resolution* of chronological evidence in both palaeoenvironmental and archaeological sequences over MIS 3 are not sufficient to confidently attribute the dramatic and abrupt climatic changes of MIS 3 as drivers of cultural change during the European Palaeolithic.

Future Work

Future studies in this field should be focused on improving model-data comparisons and assimilation in palaeoenvironmental science, improving the statistical significance of findings by increasing the resolution and numbers of measurements in palaeoenvironmental analyses, and improving the chronologies and resolution of palaeoenvironmental records in Europe. GCMs and the field of palaeoclimate modelling have significant potential for illuminating questions in archaeology and palaeoecology, but without adequate validation the conclusions drawn from their output will always be limited. Improving palaeoenvironmental chronologies by radiocarbon dating is also essential, but extremely challenging, particularly for marine sediment cores. Here, improved calibration, contaminant removal, and quantification of reservoir effects will certainly have significant impact on chronological estimates in the future.

References

- Adolphi, F., Bronk Ramsey, C., Erhardt, T., Lawrence Edwards, R., Cheng, H., Turney, C., Cooper, A., Svensson, A., Rasmussen, S. O., Fischer, H., & Muscheler, R. (2018). Connecting the Greenland ice-core and U/Th timescales via cosmogenic radionuclides: Testing the synchronicity of Dansgaard-Oeschger events. *Climate of the Past*, 14(11), 1755–1781. <https://doi.org/10.5194/cp-14-1755-2018>
- Adolphi, F., & Muscheler, R. (2016). Synchronizing the Greenland ice core and radiocarbon timescales over the Holocene-Bayesian wiggle-matching of cosmogenic radionuclide records. *Climate of the Past*, 12(1), 15–30. <https://doi.org/10.5194/cp-12-15-2016>
- Aitken, M. J. (1990). *Science-based dating in archaeology*. Longman.
- Albert, P. G., Giaccio, B., Isaia, R., Costa, A., Niespolo, E. M., Nomade, S., Pereira, A., Renne, P. R., Hinchliffe, A., Mark, D. F., Brown, R. J., & Smith, V. C. (2019). Evidence for a large-magnitude eruption from Campi Flegrei caldera (Italy) at 29 ka. *Geology*, 47(7), 595–599. <https://doi.org/10.1130/G45805.1>
- Alex, B., Barzilai, O., HersHKovitz, I., Marder, O., Berna, F., Caracuta, V., Abulafia, T., Davis, L., Goder-Goldberger, M., Lavi, R., Mintz, E., Regev, L., Mayer, D. B. Y., Tejero, J. M., Yeshurun, R., Ayalon, A., Bar-Matthews, M., Yasur, G., Frumkin, A., ... Boaretto, E. (2017). Radiocarbon chronology of Manot Cave, Israel and Upper Paleolithic dispersals. *Science Advances*, 3(11). <https://doi.org/10.1126/sciadv.1701450>
- Allen, J. R. M., Brandt, U., Brauer, A., Hubberten, H. W., Huntley, B., Keller, J., Kraml, M., Mackensen, A., Mingram, J., Negendank, J. F. W., Nowaczyk, N. R., Oberhänsli, H., Watts, W. A., Wulf, S., & Zolitschka, B. (1999). Rapid environmental changes in southern Europe during the last glacial period. *Nature*, 400(6746), 740–743. <https://doi.org/10.1038/23432>
- Allen, J. R. M., Watts, W. A., & Huntley, B. (2000). Weichselian palynostratigraphy, palaeovegetation and palaeoenvironment; the record from Lago Grande di Monticchio, southern Italy. *Quaternary International*, 73-74, 91–110. [https://doi.org/10.1016/S1165-2501\(99\)00040-0](https://doi.org/10.1016/S1165-2501(99)00040-0)

[//doi.org/10.1016/S1040-6182\(00\)00067-7](https://doi.org/10.1016/S1040-6182(00)00067-7)

- Alloway, B. V., Pribadi, A., Westgate, J. A., Bird, M. I., Fifield, L. K., Hogg, A. G., & Smith, I. (2004). Correspondence between glass-FT and ¹⁴C ages of silicic pyroclastic flow deposits sourced from Maninjau caldera, west-central Sumatra. *Earth and Planetary Science Letters*, 227(1-2), 121–133. <https://doi.org/10.1016/j.epsl.2004.08.014>
- Alon, D., Mintz, G., Cohen, I., Weiner, S., & Boaretto, E. (2002). The use of raman spectroscopy to monitor the removal of humic substances from charcoal: Quality control for ¹⁴C dating of charcoal. *Radiocarbon*, 44(1), 1–11. <https://doi.org/10.1017/S0033822200064638>
- Andel, T. van. (2005). *Neanderthals and Modern Humans in the European Landscape during the Last Glaciation: Archaeological Results of the Stage 3 Project*. McDonald Institute for Archaeological Research. <https://doi.org/10.1086/jar.61.2.3630889>
- Armstrong, E., Hopcroft, P. O., & Valdes, P. (2019). A simulated Northern Hemisphere terrestrial climate dataset for the past 60,000 years. *Scientific Data*, 6(1), 265. <https://doi.org/10.1038/s41597-019-0277-1>
- Arobba, D., Boscato, P., Boschian, G., Falgueres, C., Fasani, L., Peretto, C., Sala, B., Hohenstein, U. T., & Tozzi, C. (2004). Paleoenvironmental analysis. *Collegium Antropologicum*, 28(1), 5–21.
- Arrizabalaga, A., Iriarte-Chiapusso, M. J., & Garcia-Ibaibarriaga, N. (2021). Gravettian and Solutrean in the Basque Crossroads: Climate changes and human adaptations in the western Pyrenees. *Quaternary International*, 581-582(June 2020), 52–60. <https://doi.org/10.1016/j.quaint.2020.07.004>
- Ascough, P. L., Bird, M. I., Brock, F., Higham, T., Meredith, W., Snape, C. E., & Vane, C. H. (2009). Hydropyrolysis as a new tool for radiocarbon pre-treatment and the quantification of black carbon. *Quaternary Geochronology*, 4(2), 140–147. <https://doi.org/10.1016/j.quageo.2008.11.001>
- Ascough, P. L., Bird, M. I., Francis, S. M., & Lebl, T. (2011). Alkali extraction of archaeological and geological charcoal: Evidence for diagenetic degradation and formation of humic acids. *Journal of Archaeological Science*, 38(1), 69–78. <https://doi.org/10.1016/j.jas.2010.08.011>
- Ascough, P. L., Bird, M. I., Francis, S. M., Thornton, B., Midwood, A. J., Scott, A. C., & Apperley, D. (2011). Variability in oxidative degradation of charcoal: Influence of production conditions and environmental exposure. *Geochimica Et Cosmochimica Acta*, 75(9), 2361–2378. <https://doi.org/10.1016/j.gca.2011.02.002>

- Ascough, P. L., Bird, M. I., Meredith, W., Wood, R., Snape, C. E., Brock, F., Higham, T., Large, D. J., & Apperley, D. (2010). Hydropyrolysis: Implications for radiocarbon pretreatment and characterization of black carbon. *Radiocarbon*, 52(3), 1336–1350. <https://doi.org/10.1017/S0033822200046427>
- Ascough, P. L., Bird, M. I., Wormald, P., Snape, C. E., & Apperley, D. (2008). Influence of production variables and starting material on charcoal stable isotopic and molecular characteristics. *Geochimica Et Cosmochimica Acta*, 72(24), 6090–6102. <https://doi.org/10.1016/j.gca.2008.10.009>
- Ascough, P. L., Sturrock, C. J., & Bird, M. I. (2010). Investigation of growth responses in saprophytic fungi to charred biomass. *Isotopes in Environmental and Health Studies*, 46(1), 64–77. <https://doi.org/10.1080/10256010903388436>
- Austin, W. E. N., & Hibbert, F. D. (2012). Tracing time in the ocean: A brief review of chronological constraints (60–8 kyr) on North Atlantic marine event-based stratigraphies. *Quaternary Science Reviews*, 36, 28–37. <https://doi.org/10.1016/j.quascirev.2012.01.015>
- Azzi, C. M., Bigliocca, L., & Gulisano, F. (1977). Florence Radiocarbon Dates III. *Radiocarbon*, 19(2), 165–169. <https://doi.org/10.1017/s0033822200003507>
- Azzi, C. M., Bigliocca, L., & Piovan, E. (1973). Florence Radiocarbon Dates I. *Radiocarbon*, 15(3), 479–487. <https://doi.org/10.1017/s0033822200008948>
- Azzi, C. M., Bigliocca, L., & Piovan, E. (1974). Florence Radiocarbon Dates II. *Radiocarbon*, 16(1), 10–14. <https://doi.org/10.1017/s0033822200001405>
- Badino, F., Pini, R., Ravazzi, C., Margaritora, D., Arrighi, S., Bortolini, E., Figus, C., Giaccio, B., Lugli, F., Marciani, G., Monegato, G., Moroni, A., Negrino, F., Oxilia, G., Peresani, M., Romandini, M., Ronchitelli, A., Spinapolice, E. E., Zerboni, A., & Benazzi, S. (2020). An overview of Alpine and Mediterranean palaeogeography, terrestrial ecosystems and climate history during MIS 3 with focus on the Middle to Upper Palaeolithic transition. *Quaternary International*, 551 (January 2019), 7–28. <https://doi.org/10.1016/j.quaint.2019.09.024>
- Baillie, M. G. L., & Pilcher, J. R. (1973). A simple crossdating program for tree-ring research. *Tree Ring Bulletin*, 33, 7–14. <https://arizona.openrepository.com/arizona/handle/10150/260029>
- Banks, W. E. (2017). The application of ecological niche modeling methods to archaeological data in order to examine culture-environment relationships and cultural trajectories. *Quaternaire*, 28(2), 271–276. <https://doi.org/10.4000/quaternaire.7966>
- Banks, W. E., D’Errico, F., Dibble, H., & Krishtalka, L. (2006). Eco-cultural niche modeling: new tools for reconstructing the geography and ecology of past

- human populations. *PaleoAnthropology*, 68–83. http://www.paleoanthro.org/journal/contents%7B/_%7Ddynamic.asp?volume=2006
- Banks, W. E., D’Errico, F., & Zilhão, J. (2013). Human-climate interaction during the Early Upper Paleolithic: Testing the hypothesis of an adaptive shift between the Proto-Aurignacian and the Early Aurignacian. *Journal of Human Evolution*, 64(1), 39–55. <https://doi.org/10.1016/j.jhevol.2012.10.001>
- Baño-Medina, J., Manzananas, R., & Gutiérrez, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4), 2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>
- Barshay-Szmidt, C., Normand, C., Flas, D., & Soulier, M. C. (2018). Radiocarbon dating the Aurignacian sequence at Isturitz (France): Implications for the timing and development of the Protoaurignacian and Early Aurignacian in western Europe. *Journal of Archaeological Science: Reports*, 17(August 2017), 809–838. <https://doi.org/10.1016/j.jasrep.2017.09.003>
- Bar-Yosef, O. (2002). The Upper Paleolithic revolution. *Annual Review of Anthropology*, 31, 363–393. <https://doi.org/10.1146/annurev.anthro.31.040402.085416>
- Bataille, G., & Conard, N. J. (2018). Blade and bladelet production at Hohle Fels Cave, AH IV in the Swabian Jura and its importance for characterizing the technological variability of the Aurignacian in Central Europe. *PLoS ONE*, 13(4), e0194097. <https://doi.org/10.1371/journal.pone.0194097>
- Bay, R. C., Rohde, R. A., Price, P. B., & Bramall, N. E. (2010). South Pole paleowind from automated synthesis of ice core records. *Journal of Geophysical Research Atmospheres*, 115(14), 1–10. <https://doi.org/10.1029/2009JD013741>
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5(1), 180214. <https://doi.org/10.1038/sdata.2018.214>
- Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G., & Weatherall, P. (2009). Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. *Marine Geodesy*, 32(4), 355–371. <https://doi.org/10.1080/01490410903297766>
- Benazzi, S., Bailey, S. E., Peresani, M., Mannino, M. A., Romandini, M., Richards, M. P., & Hublin, J. J. (2014). Middle Paleolithic and Uluzzian human remains

- from Fumane cave, Italy. *Journal of Human Evolution*, 70(1), 61–68. <https://doi.org/10.1016/j.jhevol.2014.03.001>
- Benazzi, S., Douka, K., Fornai, C., Bauer, C. C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., Condemi, S., Ronchitelli, A., Harvati, K., & Weber, G. W. (2011). Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature*, 479(7374), 525–528. <https://doi.org/10.1038/nature10617>
- Benazzi, S., Slon, V., Talamo, S., Negrino, F., Peresani, M., Bailey, S. E., Sawyer, S., Panetta, D., Vicino, G., Starnini, E., Mannino, M. A., Salvadori, P. A., Meyer, M., Pääbo, S., & Hublin, J. J. (2015). The makers of the Protoaurignacian and implications for Neandertal extinction. *Science*, 348(6236), 793–796. <https://doi.org/10.1126/science.aaa2773>
- Beresford-Jones, D., Taylor, S., Paine, C., Pryor, A., Svoboda, J., & Jones, M. (2011). Rapid climate change in the Upper Palaeolithic: The record of charcoal conifer rings from the Gravettian site of Dolní Vstonice, Czech Republic. *Quaternary Science Reviews*, 30(15-16), 1948–1964. <https://doi.org/10.1016/j.quascirev.2011.04.021>
- Berto, C., Boscato, P., Boschin, F., Luzi, E., & Ronchitelli, A. (2017). Paleoenvironmental and paleoclimatic context during the Upper Palaeolithic (late Upper Pleistocene) in the Italian Peninsula. The small mammal record from Grotta Paglicci (Rignano Garganico, Foggia, Southern Italy). *Quaternary Science Reviews*, 168, 30–41. <https://doi.org/10.1016/j.quascirev.2017.05.004>
- Beyer, R., Krapp, M., Eriksson, A., & Manica, A. (2021). Climatic windows for human migration out of Africa in the past 300,000 years. *Nature Communications*, 12(1), 4889. <https://doi.org/10.1038/s41467-021-24779-1>
- Beyer, R., Krapp, M., & Manica, A. (2020). An empirical evaluation of bias correction methods for palaeoclimate simulations. *Climate of the Past*, 16(4), 1493–1508. <https://doi.org/10.5194/cp-16-1493-2020>
- Bicho, N., Marreiros, J., Cascalheira, J., Pereira, T., & Haws, J. (2015). Bayesian modeling and the chronology of the Portuguese Gravettian. *Quaternary International*, 359-360, 499–509. <https://doi.org/10.1016/j.quaint.2014.04.040>
- Bird, M. I. (2013). Charcoal. In *Encyclopedia of quaternary science: Second edition* (pp. 353–360). Elsevier. <https://doi.org/10.1016/B978-0-444-53643-3.00047-9>
- Bird, M. I., & Ascough, P. L. (2012). Isotopes in pyrogenic carbon: A review. *Organic Geochemistry*, 42(12), 1529–1539. <https://doi.org/10.1016/j.orggeochem.2010.09.005>

- Bird, M. I., Ascough, P. L., Young, I. M., Wood, C. V., & Scott, A. C. (2008). X-ray microtomographic imaging of charcoal. *Journal of Archaeological Science*, 35(10), 2698–2706. <https://doi.org/10.1016/j.jas.2008.04.018>
- Bird, M. I., Ayliffe, L. K., Fifield, L. K., Tunney, C. S. M., Cresswell, R. G., Barrows, T. T., & David, B. (1999). Radiocarbon dating of "old" charcoal using a wet oxidation, stepped-combustion procedure. *Radiocarbon*, 41(2), 127–140. <https://doi.org/10.1017/S0033822200019482>
- Bird, M. I., Charville-Mort, P. D. J., Ascough, P. L., Wood, R., Higham, T., & Apperley, D. (2010). Assessment of oxygen plasma ashing as a pre-treatment for radiocarbon dating. *Quaternary Geochronology*, 5(4), 435–442. <https://doi.org/10.1016/j.quageo.2009.10.004>
- Bird, M. I., & Gröcke, D. R. (1997). Determination of the abundance and carbon isotope composition of elemental carbon in sediments. *Geochimica Et Cosmochimica Acta*, 61(16), 3413–3423. [https://doi.org/10.1016/S0016-7037\(97\)00157-9](https://doi.org/10.1016/S0016-7037(97)00157-9)
- Bird, M. I., Levchenko, V., Ascough, P. L., Meredith, W., Wurster, C. M., Williams, A., Tilston, E. L., Snape, C. E., & Apperley, D. (2014). The efficiency of charcoal decontamination for radiocarbon dating by three pre-treatments - ABOX, ABA and hypy. *Quaternary Geochronology*, 22, 25–32. <https://doi.org/10.1016/j.quageo.2014.02.003>
- Bird, M. I., McBeath, A. V., Ascough, P. L., Levchenko, V. A., Wurster, C. M., Munksgaard, N. C., Smernik, R. J., & Williams, A. (2017). Loss and gain of carbon during char degradation. *Soil Biology and Biochemistry*, 106, 80–89. <https://doi.org/10.1016/j.soilbio.2016.12.012>
- Bird, M. I., Turney, C., Fifield, L. K., Jones, R., Ayliffe, L. K., Palmer, A., Cresswell, R. G., & Robertson, S. (2002). Radiocarbon analysis of the early archaeological site of Nauwalabila I, Arnhem Land, Australia: Implications for sample suitability and stratigraphic integrity. *Quaternary Science Reviews*, 21(8-9), 1061–1075. [https://doi.org/10.1016/S0277-3791\(01\)00058-0](https://doi.org/10.1016/S0277-3791(01)00058-0)
- Blaauw, M. (2012). Out of tune: The dangers of aligning proxy archives. *Quaternary Science Reviews*, 36, 38–49. <https://doi.org/10.1016/j.quascirev.2010.11.012>
- Blaauw, M., Wohlfarth, B., Christen, J. A., Ampel, L., Veres, D., Hughen, K. A., Preusser, F., & Svensson, A. (2010). Were last glacial climate events simultaneous between Greenland and France? A quantitative comparison using non-tuned chronologies. *Journal of Quaternary Science*, 25(3), 387–394. <https://doi.org/10.1002/jqs.1330>

- Boaretto, E., Rebollo, N. R., Cohen-Ofri, I., Popovitz-Biro, R., Bar-Yosef, O., Meignen, L., Goldberg, P., & Weiner, S. (2008). Structural characterization of charcoal exposed to high and low pH: Implications for ^{14}C sample preparation and charcoal preservation. *Radiocarbon*, 50(2), 289–307. <https://doi.org/10.1017/s0033822200033592>
- Boers, N. (2018). Early-warning signals for Dansgaard-Oeschger events in a high-resolution ice core record. *Nature Communications*, 9(1), 2556. <https://doi.org/10.1038/s41467-018-04881-7>
- Borgia, V., Boschin, F., & Ronchitelli, A. (2016). Bone and antler working at Grotta Paglicci (Rignano Garganico, Foggia, southern Italy). *Quaternary International*, 403, 23–39. <https://doi.org/10.1016/j.quaint.2015.11.116>
- Borgia, V., Ranaldo, F., Ronchitelli, A., & Wierer, U. (2011). What differences in production and use of Aurignacian and early Gravettian Lithic assemblages? The case of Grotta Paglicci (Rignano Garganico, Foggia, Southern Italy). *À La Recherche Des Identités Gravettiennes: Actualités, Questionnements Et Perspectives Acte de La Table Ronde Sur Le Gravettien En France Et Dans Les Pays Limitrophes Aix-En-Provence, 6-8 Octobre 2008.*, 161–174.
- Boscato, P. (2004). I macromammiferi dell'Aurignaziano e del Gravettiano antico di Grotta Paglicci. *Paglicci. L'Aurignaziano e Il Gravettiano Antico, January 2004*, 49–61.
- Boscato, P. (2007). Faunes gravettiennes à grands mammifères de l'Italie du Sud : Grotta della Cala (Salerno) et Grotta Paglicci (Foggia). *Paléo*, 19, 109–114. <https://doi.org/10.4000/paleo.548>
- Boscato, P., & Ronchitelli, A. (2006). La serie esterna di Paglicci. Gli scavi del 2004-2005. *XXVI Convegno Nazionale Sulla Preistoria, Protostoria e Storia Della Dauna, San Severo 10-11 Dicembre 2005*, 3–16.
- Boschin, F., Bernardini, F., Pilli, E., Vai, S., Zanolli, C., Tagliacozzo, A., Fico, R., Fedi, M., Corny, J., Dreossi, D., Lari, M., Modi, A., Vergata, C., Tuniz, C., Moroni, A., Boscato, P., Caramelli, D., & Ronchitelli, A. (2020). The first evidence for Late Pleistocene dogs in Italy. *Scientific Reports*, 10(1), 1–14. <https://doi.org/10.1038/s41598-020-69940-w>
- Boschin, F., Boscato, P., Berto, C., Crezzini, J., & Ronchitelli, A. (2018). The palaeoecological meaning of macromammal remains from archaeological sites exemplified by the case study of Grotta Paglicci (Upper Palaeolithic, southern Italy). *Quaternary Research (United States)*, 90(3), 470–482. <https://doi.org/10.1017/qua.2018.59>
- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal*,

- Source Statistical, Royal Series, Society*, 26(2), 211–252.
- Braadbaart, F., Poole, I., & Brussel, A. A. van. (2009). Preservation potential of charcoal in alkaline environments: an experimental approach and implications for the archaeological record. *Journal of Archaeological Science*, 36(8), 1672–1679. <https://doi.org/10.1016/j.jas.2009.03.006>
- Braak, C. J. F. ter, & Juggins, S. (1993). Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, 269-270(1), 485–502. <https://doi.org/10.1007/BF00028046>
- Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B. L., & Zhao, Y. (2012). Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*, 2(6), 417–424. <https://doi.org/10.1038/nclimate1456>
- Bradshaw, C. J. A., Ulm, S., Williams, A., Bird, M. I., Roberts, R. G., Jacobs, Z., Laviano, F., Weyrich, L. S., Friedrich, T., Norman, K., & Saltr , F. (2019). Minimum founding populations for the first peopling of Sahul. *Nature Ecology and Evolution*, 3(7), 1057–1063. <https://doi.org/10.1038/s41559-019-0902-6>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees* (pp. 1–358). Chapman & Hall. <https://doi.org/10.1201/9781315139470>
- Brock, F., & Higham, T. (2009). AMS radiocarbon dating of Paleolithic-aged charcoal from Europe and the Mediterranean rim using ABOX-SC. *Radiocarbon*, 51(2), 839–846. <https://doi.org/10.1017/S0033822200056149>
- Brock, F., Higham, T., Ditchfield, P., & Ramsey, C. B. (2010). Current pre-treatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon*, 52(1), 103–112. <https://doi.org/10.1017/S0033822200045069>
- Broecker, W. S., & Denton, G. H. (1990). The role of ocean-atmosphere reorganizations in glacial cycles. *Quaternary Science Reviews*, 9(4), 305–341. [https://doi.org/10.1016/0277-3791\(90\)90026-7](https://doi.org/10.1016/0277-3791(90)90026-7)
- Broglio, A., Tagliacozzo, A., De Stefani, M., Gurioli, F., & Facciolo, A. (2004). Aurignacian dwelling structures, hunting strategies and seasonality in the Fumane Cave (Lessini Mountains). *The Early Upper Paleolithic of Eurasia: General Trends, Local Developments (the Materials of International Conference Devoted to the 125th Anniversary of the Palaeolithic Investigation in Kostenki, 2004, August, 23-26)*, 2(May), 263–276.
- Budsky, A., Wassenburg, J. A., MertzKraus, R., Sp tl, C., Jochum, K. P., Gib-

- ert, L., & Scholz, D. (2019). Western Mediterranean Climate Response to Dansgaard/Oeschger Events: New Insights From Speleothem Records. *Geophysical Research Letters*, 46(15), 9042–9053. <https://doi.org/10.1029/2019GL084009>
- Bunn, A. G. (2008). A dendrochronology program library in R (dplR). *Dendrochronologia*, 26(2), 115–124. <https://doi.org/10.1016/j.dendro.2008.01.002>
- Bunn, A. G. (2010). Statistical and visual crossdating in R using the dplR library. *Dendrochronologia*, 28(4), 251–258. <https://doi.org/10.1016/j.dendro.2009.12.001>
- Buras, A. (2017). A comment on the expressed population signal. *Dendrochronologia*, 44, 130–132. <https://doi.org/10.1016/j.dendro.2017.03.005>
- Burjachs, F., López-García, J. M., Allué, E., Blain, H. A., Rivals, F., Bennàsar, M., & Expósito, I. (2012). Palaeoecology of Neanderthals during Dansgaard-Oeschger cycles in northeastern Iberia (Abric Romani): From regional to global scale. *Quaternary International*, 247(1), 26–37. <https://doi.org/10.1016/j.quaint.2011.01.035>
- Burke, A., Kageyama, M., Latombe, G., Fasel, M., Vrac, M., Ramstein, G., & James, P. M. A. (2017). Risky business: The impact of climate and climate variability on human population dynamics in Western Europe during the Last Glacial Maximum. *Quaternary Science Reviews*, 164, 217–229. <https://doi.org/10.1016/j.quascirev.2017.04.001>
- Burstyn, Y., Gazit, A., & Dvir, O. (2021). Hierarchical Dynamic Time Warping methodology for aggregating multiple geological time series. *Computers and Geosciences*, 150(December 2020), 104704. <https://doi.org/10.1016/j.cageo.2021.104704>
- Chevalier, M., Chase, B. M., Quick, L. J., Dupont, L. M., & Johnson, T. C. (2021). Temperature change in subtropical southeastern Africa during the past 790,000 yr. *Geology*, 49(1), 71–75. <https://doi.org/10.1130/G47841.1>
- Chevalier, M., Davis, B. A. S., Heiri, O., Seppä, H., Chase, B. M., Gajewski, K., Lacourse, T., Telford, R. J., Finsinger, W., Guiot, J., Kühl, N., Maezumi, S. Y., Tipton, J. R., Carter, V. A., Brussel, T., Phelps, L. N., Dawson, A., Zanon, M., Vallé, F., ... Kupriyanov, D. (2020). Pollen-based climate reconstruction techniques for late Quaternary studies. *Earth-Science Reviews*, 210(September), 103384. <https://doi.org/10.1016/j.earscirev.2020.103384>
- Cohen-Ofri, I., Popovitz-Biro, R., & Weiner, S. (2007). Structural characterization of modern and fossilized charcoal produced in natural fires as determined by

- using electron energy loss spectroscopy. *Chemistry - A European Journal*, 13(8), 2306–2310. <https://doi.org/10.1002/chem.200600920>
- Cohen-Ofri, I., Weiner, L., Boaretto, E., Mintz, G., & Weiner, S. (2006). Modern and fossil charcoal: Aspects of structure and diagenesis. *Journal of Archaeological Science*, 33(3), 428–439. <https://doi.org/10.1016/j.jas.2005.08.008>
- Columbu, A., Chiarini, V., Spötl, C., Benazzi, S., Hellstrom, J. C., Cheng, H., & De Waele, J. (2020). Speleothem record attests to stable environmental conditions during Neanderthal–modern human turnover in southern Italy. *Nature Ecology & Evolution*, 4(9), 1188–1195. <https://doi.org/10.1038/s41559-020-1243-1>
- Conard, N. J., & Bolus, M. (2003). Radiocarbon dating the appearance of modern humans and timing of cultural innovations in Europe: New results and new challenges. *Journal of Human Evolution*, 44(3), 331–371. [https://doi.org/10.1016/S0047-2484\(02\)00202-6](https://doi.org/10.1016/S0047-2484(02)00202-6)
- Constantine, M., Mooney, S., Hibbert, B., Marjo, C., Bird, M. I., Cohen, T., Forbes, M., McBeath, A., Rich, A., & Stride, J. (2021). Using charcoal, ATR FTIR and chemometrics to model the intensity of pyrolysis: Exploratory steps towards characterising fire events. *Science of the Total Environment*, 783, 147052. <https://doi.org/10.1016/j.scitotenv.2021.147052>
- Cook, E. R., Briffa, K. R., & Jones, P. D. (1994). Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *International Journal of Climatology*, 14(4), 379–402. <https://doi.org/10.1002/joc.3370140404>
- Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A., & Funkhouser, G. (1995). Chronology Development for Palaeoclimatic Studies. *The Holocene*, 5(2), 229–237.
- Cook, E. R., & Holmes, R. L. (1996). Guide for computer program ARSTAN. *The International Tree-Ring Data Bank Program Library Version*, 2(16), 75–87.
- Cook, E. R., & Peters, K. (1981). The smoothing spline, a new approach to standardising forest interior tree-ring. *Tree-Ring Bulletin*, 41, 45–53. <http://hdl.handle.net/10150/261038>
- Copenheaver, C. A., Pokorski, E. A., Currie, J. E., & Abrams, M. D. (2006). Causation of false ring formation in *Pinus banksiana*: A comparison of age, canopy class, climate and growth rate. *Forest Ecology and Management*, 236(2–3), 348–355. <https://doi.org/10.1016/j.foreco.2006.09.020>
- Corrick, E. C., Drysdale, R. N., Hellstrom, J. C., Capron, E., Rasmussen, S. O., Zhang, X., Fleitmann, D., Couchoud, I., & Wolff, E. (2020). Synchronous timing

- of abrupt climate changes during the last glacial period. *Science*, 369(6506), 963–969. <https://doi.org/10.1126/science.aay5538>
- Crezzini, J., Boscato, P., Ricci, S., Ronchitelli, A., Spagnolo, V., & Boschini, F. (2016). A spotted hyaena den in the Middle Palaeolithic of Grotta Paglicci (Gargano promontory, Apulia, Southern Italy). *Archaeological and Anthropological Sciences*, 8(2), 227–240. <https://doi.org/10.1007/s12520-015-0273-0>
- Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N. S., Hammer, C. U., Hvidberg, C. S., Steffensen, J. P., Sveinbjörnsdóttir, A. E., Jouzel, J., & Bond, G. (1993). Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, 364(6434), 218–220. <https://doi.org/10.1038/364218a0>
- De Vries, H., & Barendsen, G. W. (1954). Measurements of age by the carbon-14 technique. *Nature*, 174(4442), 1138–1141. <https://doi.org/10.1038/1741138a0>
- Díaz, S. C., Touchan, R., & Swetnam, T. W. (2001). A tree-ring reconstruction of past precipitation for Baja California Sur, Mexico. *International Journal of Climatology*, 21(8), 1007–1019. <https://doi.org/10.1002/joc.664>
- Douka, K., Chiotti, L., Nespoulet, R., & Higham, T. (2020). A refined chronology for the Gravettian sequence of Abri Pataud. *Journal of Human Evolution*, 141, 102730. <https://doi.org/10.1016/j.jhevol.2019.102730>
- Douka, K., Grimaldi, S., Boschian, G., Lucchese, A. del, & Higham, T. (2012). A new chronostratigraphic framework for the Upper Palaeolithic of Riparo Mochi (Italy). *Journal of Human Evolution*, 62(2), 286–299. <https://doi.org/10.1016/j.jhevol.2011.11.009>
- Douka, K., Higham, T., & Sinitzyn, A. (2010). The influence of pretreatment chemistry on the radiocarbon dating of Campanian Ignimbrite-aged charcoal from Kostenki 14 (Russia). *Quaternary Research*, 73(3), 583–587. <https://doi.org/10.1016/j.yqres.2010.01.002>
- Douka, K., Higham, T., Wood, R., Boscato, P., Gambassini, P., Karkanas, P., Peresani, M., & Ronchitelli, A. (2014). On the chronology of the Uluzzian. *Journal of Human Evolution*, 68(1), 1–13. <https://doi.org/10.1016/j.jhevol.2013.12.007>
- Douka, K., Slon, V., Jacobs, Z., Ramsey, C. B., Shunkov, M. V., Derevianko, A. P., Mafessoni, F., Kozlikin, M. B., Li, B., Grün, R., Comeskey, D., Deviese, T., Brown, S., Viola, B., Kinsley, L., Buckley, M., Meyer, M., Roberts, R. G., Pääbo, S., ... Higham, T. (2019). Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave. *Nature*, 565(7741), 640–644.

<https://doi.org/10.1038/s41586-018-0870-z>

- Dufourny, A., Van De Steene, L., Humbert, G., Guibal, D., Martin, L., & Blin, J. (2019). Influence of pyrolysis conditions and the nature of the wood on the quality of charcoal as a reducing agent. *Journal of Analytical and Applied Pyrolysis*, 137(October 2018), 1–13. <https://doi.org/10.1016/j.jaap.2018.10.013>
- Dyer, T. G. J., & Fritts, H. C. (1978). *Tree Rings and Climate*. (Vol. 15, p. 339). Academic Press. <https://doi.org/10.2307/2402947>
- Engelbrecht, F. A., Marean, C. W., Cowling, R. M., Engelbrecht, C. J., Neumann, F. H., Scott, L., Nkoana, R., O’Neal, D., Fisher, E. C., Shook, E., Franklin, J., Thatcher, M., McGregor, J. L., Van der Merwe, J., Dedekind, Z., & Difford, M. (2019). Downscaling Last Glacial Maximum climate over southern Africa. *Quaternary Science Reviews*, 226, 105879. <https://doi.org/10.1016/j.quascirev.2019.105879>
- Eriksson, A., Betti, L., Friend, A. D., Lycett, S. J., Singarayer, J. S., Cramon-Taubadel, N. von, Valdes, P., Balloux, F., & Manica, A. (2012). Late Pleistocene climate change and the global expansion of anatomically modern humans. *Proceedings of the National Academy of Sciences*, 109(40), 16089–16094. <https://doi.org/10.1073/pnas.1209494109>
- Evin, J., Marien, G., & Pachiaudi, C. (1979). Lyon Natural Radiocarbon Measurements VIII. *Radiocarbon*, 21(3), 405–452. <https://doi.org/10.1017/s0033822200004562>
- Falcucci, A., Conard, N. J., & Peresani, M. (2020). Breaking through the aquitaine frame: A re-evaluation on the significance of regional variants during the Aurignacian as seen from a key record in southern Europe. *Journal of Anthropological Sciences*, 98(December), 99–140. <https://doi.org/10.4436/jass.98021>
- Falcucci, A., & Peresani, M. (2019). A pre-Heinrich Event 3 assemblage at Fumane Cave and its contribution for understanding the beginning of the Gravettian in Italy. *Quartar International Yearbook for Ice Age and Stone Age Research*, 66(2019), 135–154. https://doi.org/10.7485/QU66_6
- Ferguson, C. W. (1969). A 7104-year annual tree-ring chronology for bristlecone pine, *Pinus aristata*, from the White Mountains, California. *Tree-Ring Bulletin*, 29(Schulman), 3–29.
- Fewlass, H., Talamo, S., Wacker, L., Kromer, B., Tuna, T., Fagault, Y., Bard, E., McPherron, S. P., Aldeias, V., Maria, R., Martisius, N. L., Paskulin, L., Rezek, Z., Sinet-Mathiot, V., Sirakova, S., Smith, G. M., Spasov, R., Welker, F., Sirakov, N., ... Hublin, J. J. (2020). A 14C chronology for the Middle to Upper Palaeolithic transition at Bacho Kiro Cave, Bulgaria. *Nature Ecology and*

- Evolution*, 4(6), 794–801. <https://doi.org/10.1038/s41559-020-1136-3>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Fletcher, W. J., Sánchez Goñi, M. F., Allen, J. R. M., Cheddadi, R., Combourieu-Nebout, N., Huntley, B., Lawson, I. T., Londeix, L., Magri, D., Margari, V., Müller, U. C., Naughton, F., Novenko, E., Roucoux, K., & Tzedakis, P. C. (2010). Millennial-scale variability during the last glacial in vegetation records from Europe. *Quaternary Science Reviews*, 29(21-22), 2839–2864. <https://doi.org/10.1016/j.quascirev.2009.11.015>
- Folgado, D., Barandas, M., Matias, R., Martins, R., Carvalho, M., & Gamboa, H. (2018). Time Alignment Measurement for Time Series. *Pattern Recognition*, 81, 268–279. <https://doi.org/10.1016/j.patcog.2018.04.003>
- Follieri, M., Giardini, M., Magri, D., & Sadori, L. (1998). Palynostratigraphy of the Last Glacial period in the volcanic region of central Italy. *Quaternary International*, 47-48(97), 3–20. [https://doi.org/10.1016/S1040-6182\(97\)00065-7](https://doi.org/10.1016/S1040-6182(97)00065-7)
- Franklin, J., Potts, A. J., Fisher, E. C., Cowling, R. M., & Marean, C. W. (2015). Paleodistribution modeling in archaeology and paleoanthropology. *Quaternary Science Reviews*, 110, 1–14. <https://doi.org/10.1016/j.quascirev.2014.12.015>
- Friedrich, M., Remmele, S., Kromer, B., Hofmann, J., Spurk, M., Kaiser, K. F., Orsel, C., & Küppers, M. (2004). The 12,460-year Hohenheim oak and pine tree-ring chronology from Central Europe; A unique annual record for radiocarbon calibration and paleoenvironment reconstructions. *Radiocarbon*, 46(3), 1111–1122. <https://doi.org/10.1017/S003382220003304X>
- Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., Viola, B., Prüfer, K., Meyer, M., Kelso, J., Reich, D., & Pääbo, S. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, 524(7564), 216–219. <https://doi.org/10.1038/nature14558>
- García-González, I. (2008). Comparison of different distance measures for cluster analysis of tree-ring series. *Tree-Ring Research*, 64(1), 27–37. <https://doi.org/10.3959/2007-2.1>
- Genty, D., Combourieu-Nebout, N., Peyron, O., Blamart, D., Wainer, K., Mansuri, F., Ghaleb, B., Isabello, L., Dormoy, I., Grafenstein, U. von, Bonelli, S., Landais, A., & Brauer, A. (2010). Isotopic characterization of rapid climatic events during OIS3 and OIS4 in Villars Cave stalagmites (SW-France) and correlation with

- Atlantic and Mediterranean pollen records. *Quaternary Science Reviews*, 29(19–20), 2799–2820. <https://doi.org/10.1016/j.quascirev.2010.06.035>
- Giaccio, B., Hajdas, I., Peresani, M., Fedele, F. G., & R., I. (2006). *The Campanian Ignimbrite (c. 40 ka BP) and its relevance for the timing of the Middle to Upper Palaeolithic shift: timescales and regional correlations* (pp. 343–375).
- Giaccio, B., Isaia, R., Fedele, F. G., Di Canzio, E., Hoffecker, J. F., Ronchitelli, A., Sinitsyn, A., Anikovich, M., Lisitsyn, S. N., & Popov, V. V. (2008). The Campanian Ignimbrite and Codola tephra layers: Two temporal/stratigraphic markers for the Early Upper Palaeolithic in southern Italy and eastern Europe. *Journal of Volcanology and Geothermal Research*, 177(1), 208–226. <https://doi.org/10.1016/j.jvolgeores.2007.10.007>
- Giampoudakis, K., Marske, K. A., Borregaard, M. K., Ugan, A., Singarayer, J. S., Valdes, P., Rahbek, C., & NoguésBravo, D. (2017). Niche dynamics of Palaeolithic modern humans during the settlement of the Palaearctic. *Global Ecology and Biogeography*, 26(3), 359–370. <https://doi.org/10.1111/geb.12543>
- Gillespie, R. (1997). Burnt and unburnt carbon: Dating charcoal and burnt bone from the Willandra Lakes, Australia. *Radiocarbon*, 39(3), 239–250. <https://doi.org/10.1017/S0033822200053236>
- Gillespie, R. (2019). A novel cellulose-preparation method. *Radiocarbon*, 61(1), 131–139. <https://doi.org/10.1017/RDC.2018.65>
- Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software*, 31(7), 1–24. <https://doi.org/10.18637/jss.v031.i07>
- Goring-Morris, A. N., & Belfer-Cohen, A. (2020). Noisy beginnings: The Initial Upper Palaeolithic in Southwest Asia. *Quaternary International*, 551(December 2019), 40–46. <https://doi.org/10.1016/j.quaint.2020.01.017>
- Grant, K. M., Rohling, E. J., Bar-Matthews, M., Ayalon, A., Medina-Elizalde, M., Ramsey, C. B., Satow, C., & Roberts, A. P. (2012). Rapid coupling between ice volume and polar temperature over the past 50,000 years. *Nature*, 491(7426), 744–747. <https://doi.org/10.1038/nature11593>
- Grissino-Mayer, H. (2001). Evaluating crossdating accuracy: A Manual and Tutorial for the computer program COFECHA. *Tree-Ring Research*, 57(2), 205–221. <https://repository.arizona.edu/handle/10150/251654>
- Grundmann, B. M., Bonn, S., & Roloff, A. (2008). Cross-dating of highly sensitive Common beech (*Fagus sylvatica* L.) tree-ring series with numerous missing rings. *Dendrochronologia*, 26(2), 109–113. <https://doi.org/10.1016/j.dendro.2008.05.002>

- Haesaerts, P. (2013). ABA and ABOx Radiocarbon Cross-Dating on Charcoal from Middle Pleniglacial Loess Deposits in Austria, Moravia, and Western Ukraine. *Radiocarbon*, 55(3–4), 641–647. https://doi.org/10.2458/azu_js_rc.55.16344
- Hagen, C. J., & Harper, J. T. (2023). Dynamic time warping to quantify age distortion in firn cores impacted by melt processes. *Annals of Glaciology*, 1–9. <https://doi.org/10.1017/aog.2023.52>
- Haig, J., Ascough, P. L., Wurster, C. M., & Bird, M. I. (2020). A rapid throughput technique to isolate pyrogenic carbon by hydrogen pyrolysis for stable isotope and radiocarbon analysis. *Rapid Communications in Mass Spectrometry*, 34(10), 1–11. <https://doi.org/10.1002/rcm.8737>
- Hajdas, I., Ascough, P. L., Garnett, M. H., Fallon, S. J., Pearson, C. L., Quarta, G., Spalding, K. L., Yamaguchi, H., & Yoneda, M. (2021). Radiocarbon dating. *Nature Reviews Methods Primers*, 1(1). <https://doi.org/10.1038/s43586-021-00058-7>
- Hajdas, I., Bonani, G., Thut, J., Leone, G., Pfenninger, R., & Maden, C. (2004). A report on sample preparation at the ETH/PSI AMS facility in Zurich. *Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms*, 223–224(SPEC. ISS.), 267–271. <https://doi.org/10.1016/j.nimb.2004.04.054>
- Harrison, S. P., Prentice, I. C., Barboni, D., Kohfeld, K. E., Ni, J., & Sutra, J. P. (2010). Ecophysiological and bioclimatic foundations for a global plant functional classification. *Journal of Vegetation Science*, 21(2), 300–317. <https://doi.org/10.1111/j.1654-1103.2009.01144.x>
- Harrison, S. P., & Sanchez Goñi, M. F. (2010). Global patterns of vegetation response to millennial-scale variability and rapid climate change during the last glacial period. *Quaternary Science Reviews*, 29(21–22), 2957–2980. <https://doi.org/10.1016/j.quascirev.2010.07.016>
- Hassan, M. M., Jones, E., & Buck, C. E. (2019). A simple Bayesian approach to tree-ring dating. *Archaeometry*, 61(4), 991–1010. <https://doi.org/10.1111/arcm.12466>
- Hatté, C., Morvan, J., Noury, C., & Paterne, M. (2001). Is classical acid-alkali-acid treatment responsible for contamination? An alternative proposition. *Radiocarbon*, 43(2 PART I), 177–182. <https://doi.org/10.1017/s003382220003798x>
- Hauck, T. C., Lehmkuhl, F., Zeeden, C., Böskén, J., Thiemann, A., & Richter, J. (2018). The Aurignacian way of life: Contextualizing early modern human adaptation in the Carpathian Basin. *Quaternary International*, 485, 150–166.

<https://doi.org/10.1016/j.quaint.2017.10.020>

- Haws, J., Benedetti, M., Carvalho, M., Ellis, G., Pereira, T., Cascalheira, J., Bicho, N., & Friedl, L. (2021). Human adaptive responses to climate and environmental change during the Gravettian of Lapa do Picareiro (Portugal). *Quaternary International*, 587-588(April 2020), 4–18. <https://doi.org/10.1016/j.quaint.2020.08.009>
- Haws, J., Benedetti, M., Talamo, S., Bicho, N., Cascalheira, J., Ellis, G., Carvalho, M., Friedl, L., Pereira, T., & Zinsious, B. K. (2020). The early Aurignacian dispersal of modern humans into westernmost Eurasia. *Proceedings of the National Academy of Sciences*, 117(41), 25414–25422. <https://doi.org/10.1073/pnas.2016062117>
- Hay, C. C., Creveling, J. R., Hagen, C. J., Maloof, A. C., & Huybers, P. (2019). A library of early Cambrian chemostratigraphic correlations from a reproducible algorithm. *Geology*, 47(5), 457–460. <https://doi.org/10.1130/G46019.1>
- Haywood, A. M., Valdes, P., Aze, T., Barlow, N., Burke, A., Dolan, A. M., Heydt, A. S. von der, Hill, D. J., Jamieson, S. S. R., Otto-Bliesner, B. L., Salzmann, U., Saupe, E., & Voss, J. (2019). What can Palaeoclimate Modelling do for you? *Earth Systems and Environment*, 3(1), 1–18. <https://doi.org/10.1007/s41748-019-00093-1>
- Head, M. J. (1987). Categorisation of organic sediments from archaeological sites. In W. Ambrose & J. Mummary (Eds.), *Archaeometry: Further australasian studies* (pp. 143–169). Research School of Pacific Studies, ANU.
- Henry, L. G., McManus, J. F., Curry, W. B., Roberts, N. L., Piotrowski, A. M., & Keigwin, L. D. (2016). North Atlantic ocean circulation and abrupt climate change during the last glaciation. *Science*, 353(6298), 470–474. <https://doi.org/10.1126/science.aaf5529>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., MuñozSabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Higham, T. (2011). European middle and upper palaeolithic radiocarbon dates are often older than they look: Problems with previous dates and some remedies. *Antiquity*, 85(327), 235–249. <https://doi.org/10.1017/S0003598X00067570>
- Higham, T., Barton, H., Turney, C., Barker, G., Ramsey, C. B., & Brock, F. (2009). Radiocarbon dating of charcoal from tropical sequences: Results from the Niah

- Great Cave, Sarawak, and their broader implications. *Journal of Quaternary Science*, 24(2), 189–197. <https://doi.org/10.1002/jqs.1197>
- Higham, T., Basell, L., Jacobi, R., Wood, R., Ramsey, C. B., & Conard, N. J. (2012). Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: The radiocarbon chronology of GeiSSenklösterle. *Journal of Human Evolution*, 62(6), 664–676. <https://doi.org/10.1016/j.jhevol.2012.03.003>
- Higham, T., Brock, F., Peresani, M., Broglio, A., Wood, R., & Douka, K. (2009). Problems with radiocarbon dating the Middle to Upper Palaeolithic transition in Italy. *Quaternary Science Reviews*, 28(13-14), 1257–1267. <https://doi.org/10.1016/j.quascirev.2008.12.018>
- Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruiz, C., Bergman, C., Boitard, C., Boscato, P., Caparrós, M., Conard, N. J., Draily, C., Froment, A., Galván, B., Gambassini, P., ... Jacobi, R. (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*, 512(7514), 306–309. <https://doi.org/10.1038/nature13621>
- Higham, T., Jacobi, R., Basell, L., Ramsey, C. B., Chiotti, L., & Nespoulet, R. (2011). Precision dating of the Palaeolithic: A new radiocarbon chronology for the Abri Pataud (France), a key Aurignacian sequence. *Journal of Human Evolution*, 61(5), 549–563. <https://doi.org/10.1016/j.jhevol.2011.06.005>
- Higham, T., Jacobi, R., Julien, M., David, F., Basell, L., Wood, R., Davies, W., & Ramsey, C. B. (2010). Chronology of the Grotte du Renne (France) and implications for the context of ornaments and human remains within the Châtelperronian. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47), 20234–20239. <https://doi.org/10.1073/pnas.1007963107>
- Hoffecker, J. F. (2005). Innovation and technological knowledge in the Upper Paleolithic of Northern Eurasia. *Evolutionary Anthropology*, 14(5), 186–198. <https://doi.org/10.1002/evan.20066>
- Hoffecker, J. F. (2009). The spread of modern humans in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38), 16040–16045. <https://doi.org/10.1073/pnas.0903446106>
- Hogg, A. G., Southon, J., Turney, C., Palmer, J. G., Bronk Ramsey, C., Fenwick, P., Boswijk, G., Friedrich, M., Helle, G., Hughen, K. A., Jones, R., Kromer, B., Noronha, A., Reynard, L., Staff, R., & Wacker, L. (2016). Punctuated Shutdown of Atlantic Meridional Overturning Circulation during Greenland Stadial 1.

- Scientific Reports*, 6(May), 1–10. <https://doi.org/10.1038/srep25902>
- Holmes, R. L. (1983). Computer-Assisted Quality Control in Tree-Ring Dating and Measurement. *Tree-Ring Bulletin*, 43, 69–78.
- Intergovernmental Panel on Climate Change. (2014). *Climate Change 2013 – The Physical Science Basis* (Intergovernmental Panel on Climate Change, Ed.; p. 104). Cambridge University Press. <https://doi.org/10.1017/CB09781107415324>
- Intergovernmental Panel on Climate Change. (2023). *Climate Change 2021 – The Physical Science Basis*. Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- Izumi, K., Armstrong, E., & Valdes, P. (2023). Global footprints of dansgaard-oeschger oscillations in a GCM. *Quaternary Science Reviews*, 305, 108016. <https://doi.org/10.1016/j.quascirev.2023.108016>
- Jennings, R. P., Singarayer, J. S., Stone, E. J., Krebs-Kanzow, U., Khon, V., Nisancioglu, K. H., Pfeiffer, M., Zhang, X., Parker, A., Parton, A., Groucutt, H. S., White, T. S., Drake, N. A., & Petraglia, M. D. (2015). The greening of Arabia: Multiple opportunities for human occupation of the Arabian Peninsula during the Late Pleistocene inferred from an ensemble of climate model simulations. *Quaternary International*, 382, 181–199. <https://doi.org/10.1016/j.quaint.2015.01.006>
- Jones, T. D., Lawson, I. T., Reed, J. M., Wilson, G. P., Leng, M. J., Gierga, M., Bernasconi, S. M., Smittenberg, R. H., Hajdas, I., Bryant, C. L., & Tzedakis, P. C. (2013). Diatom-inferred late Pleistocene and Holocene palaeolimnological changes in the Ioannina basin, northwest Greece. *Journal of Paleolimnology*, 49(2), 185–204. <https://doi.org/10.1007/s10933-012-9654-x>
- Katz, R. (2002). Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research*, 20(2), 167–185. <https://doi.org/10.3354/cr020167>
- Kim, J. C., Park, J. H., Kim, I. C., Lee, C., Cheoun, M. K., Kang, J., & Song, Y. M. (2001). Progress at the Seoul National University AMS Facility. *Radiocarbon*, 43(2A), 163–167. <https://doi.org/10.1017/S0033822200037966>
- Kindler, P., Guillevic, M., Baumgartner, M., Schwander, J., Landais, A., & Leuenberger, M. (2014). Temperature reconstruction from 10 to 120 kyr b2k from the NGRIP ice core. *Climate of the Past*, 10(2), 887–902. <https://doi.org/10.5194/cp-10-887-2014>
- Kozowski, J. K. (2015). The origin of the Gravettian. *Quaternary International*, 359–360, 3–18. <https://doi.org/10.1016/j.quaint.2014.03.025>

- Krapp, M., Beyer, R., Edmundson, S. L., Valdes, P., & Manica, A. (2021). A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years. *Scientific Data*, 8(1), 228. <https://doi.org/10.1038/s41597-021-01009-3>
- Kruskal, J. B. (1983). An Overview of Sequence Comparison: Time Warps, String Edits, and Macromolecules. *SIAM Review*, 25(2), 201–237. <https://doi.org/10.1137/1025045>
- Kühl, N., Gebhardt, C., Litt, T., & Hense, A. (2002). Probability density functions as botanical-climatological transfer functions for climate reconstruction. *Quaternary Research*, 58(3), 381–392. <https://doi.org/10.1006/qres.2002.2380>
- Kuhn, M. (2016). *caret: Classification and Regression Training*. R package version 6.0-68. <https://cran.r-project.org/package=caret>.
- Kuhn, S. L. (2019). Initial Upper Paleolithic: A (near) global problem and a global opportunity. *Archaeological Research in Asia*, 17(October 2017), 2–8. <https://doi.org/10.1016/j.ara.2018.10.002>
- Kuhn, S. L., & Zwyns, N. (2014). Rethinking the initial Upper Paleolithic. *Quaternary International*, 347(1), 29–38. <https://doi.org/10.1016/j.quaint.2014.05.040>
- Larrasoana, J. C., Roberts, A. P., & Rohling, E. J. (2013). Dynamics of Green Sahara Periods and Their Role in Hominin Evolution. *PLoS ONE*, 8(10), e76514. <https://doi.org/10.1371/journal.pone.0076514>
- Latecki, L. J., Megalooikonomou, V., Wang, Q., & Yu, D. (2007). An elastic partial shape matching technique. *Pattern Recognition*, 40(11), 3069–3080. <https://doi.org/10.1016/j.patcog.2007.03.004>
- Latombe, G., Burke, A., Vrac, M., Levavasseur, G., Dumas, C., Kageyama, M., & Ramstein, G. (2018). Comparison of spatial downscaling methods of general circulation model results to study climate variability during the Last Glacial Maximum. *Geoscientific Model Development*, 11(7), 2563–2579. <https://doi.org/10.5194/gmd-11-2563-2018>
- Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*, 528(7580), 60–68. <https://doi.org/10.1038/nature16069>
- Levavasseur, G., Vrac, M., Roche, D. M., Paillard, D., Martin, A., & Vandenberghe, J. (2011). Present and LGM permafrost from climate simulations: contribution of statistical downscaling. *Climate of the Past*, 7(4), 1225–1246. <https://doi.org/10.5194/cp-7-1225-2011>
- Li, D. (2022). If a tree grows no ring and no one is around: how scientists deal with missing tree rings. *Climatic Change*, 174(1-2), 6. <https://doi.org/10>

.1007/s10584-022-03424-w

- Li, J., Deng, L., Haeb-Umbach, R., & Gong, Y. (2016). Fundamentals of speech recognition. In *Robust automatic speech recognition* (pp. 9–40). Elsevier. <https://doi.org/10.1016/b978-0-12-802398-3.00002-7>
- Lin, L., Khider, D., Lisiecki, L. E., & Lawrence, C. E. (2014). Probabilistic sequence alignment of stratigraphic records. *Paleoceanography*, 29(10), 976–989. <https://doi.org/10.1002/2014PA002713>
- Lippi, M. M., Foggi, B., Aranguren, B., Ronchitelli, A., & Revedin, A. (2015). Multistep food plant processing at Grotta Paglicci (Southern Italy) around 32,600 cal B.P. *Proceedings of the National Academy of Sciences of the United States of America*, 112(39), 12075–12080. <https://doi.org/10.1073/pnas.1505213112>
- Lisiecki, L. E., & Lisiecki, P. A. (2002). Application of dynamic programming to the correlation of paleoclimate records. *Paleoceanography*, 17(4), 1-1-1-12. <https://doi.org/10.1029/2001pa000733>
- Loader, N. J., Mccarroll, D., Miles, D., Young, G. H. F., Davies, D., & Ramsey, C. B. (2019). Tree ring dating using oxygen isotopes: a master chronology for central England. *Journal of Quaternary Science*, 34(6), 475–490. <https://doi.org/10.1002/jqs.3115>
- Loftus, E., Pargeter, J., Mackay, A., Stewart, B., & Mitchell, P. (2019). Late Pleistocene human occupation in the Maloti-Drakensberg region of southern Africa: New radiocarbon dates from Rose Cottage Cave and inter-site comparisons. *Journal of Anthropological Archaeology*, 56(November), 101117. <https://doi.org/10.1016/j.jaa.2019.101117>
- López-García, J. M., Berto, C., Luzi, E., Valle, C. D., Bañuls-Cardona, S., & Sala, B. (2015). The genus *Iberomys* (CHALINE, 1972) (Rodentia, Arvicolinae, Mammalia) in the Pleistocene of Italy. *Italian Journal of Geosciences*, 134(1), 162–169. <https://doi.org/10.3301/IJG.2014.48>
- Lorrey, A. M., Boswijk, G., Hogg, A. G., Palmer, J. G., Turney, C., Fowler, A. M., Ogden, J., & Woolley, J. M. (2018). The scientific value and potential of New Zealand swamp kauri. *Quaternary Science Reviews*, 183, 124–139. <https://doi.org/10.1016/j.quascirev.2017.12.019>
- Lougheed, B. C., Filipsson, H. L., & Snowball, I. (2013). Large spatial variations in coastal ^{14}C reservoir age - A case study from the Baltic Sea. *Climate of the Past*, 9(3), 1015–1028. <https://doi.org/10.5194/cp-9-1015-2013>
- Lougheed, B. C., & Obrochta, S. P. (2019). A Rapid, Deterministic Age-Depth Modeling Routine for Geological Sequences With Inherent Depth Uncertainty.

- Paleoceanography and Paleoclimatology*, 34(1), 122–133. <https://doi.org/10.1029/2018PA003457>
- Magri, D. (1999). Late Quaternary vegetation history at Lagaccione near Lago di Bolsena (central Italy). *Review of Palaeobotany and Palynology*, 106(3–4), 171–208. [https://doi.org/10.1016/S0034-6667\(99\)00006-8](https://doi.org/10.1016/S0034-6667(99)00006-8)
- Maier, A. (2017). Population and settlement dynamics from the Gravettian to the Magdalenian. *Mitteilungen Der Gesellschaft für Urgeschichte*, 26, 83–101. <https://www.researchgate.net/publication/323476365>
- Malmierca-Vallet, I., & Sime, L. C. (2023). Dansgaard–Oeschger events in climate models: review and baseline Marine Isotope Stage 3 (MIS3) protocol. *Climate of the Past*, 19(5), 915–942. <https://doi.org/10.5194/cp-19-915-2023>
- Maraun, D., & Widmann, M. (2018). *Statistical Downscaling and Bias Correction for Climate Research*. <https://doi.org/10.1017/9781107588783>
- Marciani, G., Ronchitelli, A., Arrighi, S., Badino, F., Bortolini, E., Boscato, P., Boschini, F., Crezzini, J., Delpiano, D., Falcucci, A., Figus, C., Lugli, F., Oxilia, G., Romandini, M., Riel-Salvatore, J., Negrino, F., Peresani, M., Spinapolice, E. E., Moroni, A., & Benazzi, S. (2020). Lithic techno-complexes in Italy from 50 to 39 thousand years BP: An overview of lithic technological changes across the Middle-Upper Palaeolithic boundary. *Quaternary International*, 551 (November 2019), 123–149. <https://doi.org/10.1016/j.quaint.2019.11.005>
- Margari, V., Pyle, D. M., Bryant, C. L., & Gibbard, P. L. (2007). Mediterranean tephra stratigraphy revisited: Results from a long terrestrial sequence on Lesbos Island, Greece. *Journal of Volcanology and Geothermal Research*, 163(1–4), 34–54. <https://doi.org/10.1016/j.jvolgeores.2007.02.002>
- Marín-Arroyo, A. B., Rios-Garaizar, J., Straus, L. G., Jones, J. R., De la Rasilla, M., González Morales, M. R., Richards, M. P., Altuna, J., Mariezkurrena, K., & Ocío, D. (2018). Chronological reassessment of the Middle to Upper Paleolithic transition and Early Upper Paleolithic cultures in Cantabrian Spain. *PLoS ONE*, 13(4), 1–20. <https://doi.org/10.1371/journal.pone.0194708>
- Marín-Arroyo, A. B., Terlato, G., Vidal-Cordasco, M., & Peresani, M. (2023). Subsistence of early anatomically modern humans in Europe as evidenced in the Protoaurignacian occupations of Fumane Cave, Italy. *Scientific Reports*, 13(1), 3788. <https://doi.org/10.1038/s41598-023-30059-3>
- Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L., & Brewer, S. (2018). Reconciling divergent trends and millennial variations in Holocene temperatures. *Nature*, 554(7690), 92–96. <https://doi.org/10.1038/nature25464>
- Mayer, J. H., Burr, G. S., & Holliday, V. T. (2008). Comparisons and interpre-

- tations of charcoal and organic matter radiocarbon ages from buried soils in North Central Colorado, USA. *Radiocarbon*, 50(3), 331–346. <https://doi.org/10.1017/S0033822200053479>
- McCarroll, D., & Loader, N. J. (2004). Stable isotopes in tree rings. *Quaternary Science Reviews*, 23(7-8), 771–801. <https://doi.org/10.1016/j.quascirev.2003.06.017>
- McGuire, A. M., Lane, C. S., Roucoux, K. H., Albert, P., & Kearney, R. (2022). The dating and correlation of an eastern Mediterranean lake sediment sequence: a 46–4 ka tepthrostratigraphy for Ioannina (NW Greece). *Journal of Quaternary Science*, 37(8), 1313–1331. <https://doi.org/10.1002/jqs.3452>
- Mellars, P. (2006). A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature*, 439(7079), 931–935. <https://doi.org/10.1038/nature04521>
- Melvin, T. M., & Briffa, K. R. (2008). A "signal-free" approach to dendroclimatic standardisation. *Dendrochronologia*, 26(2), 71–86. <https://doi.org/10.1016/j.dendro.2007.12.001>
- Menviel, L. C., Skinner, L. C., Tarasov, L., & Tzedakis, P. C. (2020). An ice–climate oscillatory framework for Dansgaard–Oeschger cycles. *Nature Reviews Earth & Environment*, 1(12), 677–693. <https://doi.org/10.1038/s43017-020-00106-y>
- Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411(September), 108815. <https://doi.org/10.1016/j.ecolmodel.2019.108815>
- Moseley, G. E., Spötl, C., Svensson, A. M., Cheng, H., Brandstätter, S., & Edwards, R. L. (2014). Multi-speleothem record reveals tightly coupled climate between central europe and greenland during marine isotope stage 3. *Geology*, 42(12), 1043–1046. <https://doi.org/10.1130/G36063.1>
- Muglia, J., Mulitza, S., Repschläger, J., Schmittner, A., Lembke-Jene, L., Lisiecki, L., Mix, A., Saraswat, R., Sikes, E., Waelbroeck, C., Gottschalk, J., Lippold, J., Lund, D., Martinez-Mendez, G., Michel, E., Muschitiello, F., Naik, S., Okazaki, Y., Stott, L., ... Zhao, N. (2023). A global synthesis of high-resolution stable isotope data from benthic foraminifera of the last deglaciation. *Scientific Data*, 10(1), 1–23. <https://doi.org/10.1038/s41597-023-02024-2>
- Müller, U. C., Pross, J., Tzedakis, P. C., Gamble, C. S., Kotthoff, U., Schmiedl, G., Wulf, S., & Christanis, K. (2011). The role of climate in the spread of modern humans into Europe. *Quaternary Science Reviews*, 30(3-4), 273–279.

<https://doi.org/10.1016/j.quascirev.2010.11.016>

- Muschitiello, F., & Aquino-lopez, M. A. (2023). *Continuous synchronization of the Greenland ice-core and U-Th timescales using probabilistic inversion*. 2005(August), 1–32.
- Muschitiello, F., O'Regan, M., Martens, J., West, G., Gustafsson, Ö., & Jakobsson, M. (2020). A new 30 000-year chronology for rapidly deposited sediments on the Lomonosov Ridge using bulk radiocarbon dating and probabilistic stratigraphic alignment. *Geochronology*, 2(1), 81–91. <https://doi.org/10.5194/gchron-2-81-2020>
- New, M., Lister, D., Hulme, M., & Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate Research*, 21(1), 1–25. <https://doi.org/10.3354/cr021001>
- Nigst, P. R., Haesaerts, P., Damblon, F., Frank-Fellner, C., Mallol, C., Viola, B., Göttinger, M., Niven, L., Trnka, G., & Hublin, J. J. (2014). Early modern human settlement of Europe north of the Alps occurred 43,500 years ago in a cold steppe-type environment. *Proceedings of the National Academy of Sciences of the United States of America*, 111(40), 14394–14399. <https://doi.org/10.1073/pnas.1412201111>
- Oriani, F., Treble, P. C., Baker, A., & Mariethoz, G. (2022). WlCount: Geological lamination detection and counting using an image analysis approach. *Computers and Geosciences*, 160(December 2021), 105037. <https://doi.org/10.1016/j.cageo.2022.105037>
- Orr, T. J., Wurster, C. M., Levchenko, V., Ascough, P. L., & Bird, M. I. (2021). Improved pretreatment method for the isolation and decontamination of pyrogenic carbon for radiocarbon dating using hydrogen pyrolysis. *Quaternary Geochronology*, 61(February 2020), 101124. <https://doi.org/10.1016/j.quageo.2020.101124>
- Palma di Cesnola, A. (2004). *Paglicci: L'Aurignanziano e il Gravettiano antico* (A. Palma di Cesnola, Ed.). Claudio Grenzi.
- Palma di Cesnola, A. (2006). L'Aurignacien et le Gravettien ancien de la grotte Paglicci au Mont Gargano. *Anthropologie*, 110(3), 355–370. <https://doi.org/10.1016/j.anthro.2006.06.011>
- Palmer, J. G., Turney, C., Fenwick, P., Thomas, Z., Helle, G., Jones, R., Clement, A., Hogg, A. G., Southon, J., Bronk Ramsey, C., Staff, R., Muscheler, R., Corrège, T., & Hua, Q. (2016). Changes in El Niño – Southern Oscillation (ENSO) conditions during the Greenland Stadial 1 (GS-1) chronozone revealed by New Zealand tree-rings. *Quaternary Science Reviews*, 153, 139–155. <https://doi.org/10.1016/j.quascirev.2016.08.011>

[//doi.org/10.1016/j.quascirev.2016.10.003](https://doi.org/10.1016/j.quascirev.2016.10.003)

- Panyushkina, I. P., Chang, C., Clement, A., & Bykov, N. (2010). First tree-ring chronology from Andronovo archaeological timbers of Bronze Age in Central Asia. *Dendrochronologia*, 28(1), 13–21. <https://doi.org/10.1016/j.dendro.2008.10.001>
- Pargeter, J., Loftus, E., Mackay, A., Mitchell, P., & Stewart, B. (2018). New ages from Boomplaas Cave, South Africa, provide increased resolution on late/terminal Pleistocene human behavioural variability. *Azania*, 53(2), 156–184. <https://doi.org/10.1080/0067270X.2018.1436740>
- Paterne, M., Kallel, N., Labeyrie, L., Vautravers, M., Duplessy, J. C., Rossignol-Strick, M., Cortijo, E., Arnold, M., & Fontugne, M. (1999). Hydrological relationship between the North Atlantic Ocean and the Mediterranean Sea during the past 15 - 75 kyr. *Paleoceanography*, 14(5), 626–638. <https://doi.org/10.1029/1998PA900022>
- Pearson, C. L., Leavitt, S. W., Kromer, B., Solanki, S. K., & Usoskin, I. (2022). Dendrochronology and Radiocarbon Dating. *Radiocarbon*, 64(3), 569–588. <https://doi.org/10.1017/RDC.2021.97>
- Pederzani, S., Britton, K., Aldeias, V., Bourgon, N., Fewlass, H., Lauer, T., McPherron, S. P., Rezek, Z., Sirakov, N., Smith, G. M., Spasov, R., Tran, N. H., Tsanova, T., & Hublin, J. J. (2021). Subarctic climate for the earliest Homo sapiens in Europe. *Science Advances*, 7(39). <https://doi.org/10.1126/sciadv.abi4642>
- Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5), 1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Peresani, M., Cremaschi, M., Ferraro, F., Falgueres, C., Bahain, J. J., Gruppioni, G., Sibilia, E., Quarta, G., Calcagnile, L., & Dolo, J. M. (2008). Age of the final Middle Palaeolithic and Uluzzian levels at Fumane Cave, Northern Italy, using ¹⁴C, ESR, ²³⁴U/²³⁰Th and thermoluminescence methods. *Journal of Archaeological Science*, 35(11), 2986–2996. <https://doi.org/10.1016/j.jas.2008.06.013>
- Peresani, M., Cristiani, E., & Romandini, M. (2016). The Uluzzian technology of Grotta di Fumane and its implication for reconstructing cultural dynamics in the Middle-Upper Palaeolithic transition of Western Eurasia. *Journal of Human Evolution*, 91, 36–56. <https://doi.org/10.1016/j.jhevol.2015.10.012>
- Peresani, M., Romandini, M., Duches, R., Jéquier, C., Nannini, N., Pastoors, A., Picin, A., Schmidt, I., Vaquero, M., & Weniger, G. C. (2014). New evidence

- for the Mousterian and Gravettian at Rio Secco Cave, Italy. *Journal of Field Archaeology*, 39(4), 401–416. <https://doi.org/10.1179/0093469014Z.0000000098>
- Pérez-Mejías, C., Moreno, A., Sancho, C., Martín-García, R., Spötl, C., Cacho, I., Cheng, H., & Edwards, R. L. (2019). Orbital-to-millennial scale climate variability during Marine Isotope Stages 5 to 3 in northeast Iberia. *Quaternary Science Reviews*, 224. <https://doi.org/10.1016/j.quascirev.2019.105946>
- Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3), 678–693. <https://doi.org/10.1016/j.patcog.2010.09.013>
- Pettitt, P. B., Davies, W., Gamble, C. S., & Richards, M. B. (2003). Palaeolithic radiocarbon chronology: Quantifying our confidence beyond two half-lives. *Journal of Archaeological Science*, 30(12), 1685–1693. [https://doi.org/10.1016/S0305-4403\(03\)00070-0](https://doi.org/10.1016/S0305-4403(03)00070-0)
- Pietikäinen, J., Kiikkilä, O., & Fritze, H. (2000). Charcoal as a habitat for microbes and its effect on the microbial community of the underlying humus. *Oikos*, 89(2), 231–242. <https://doi.org/10.1034/j.1600-0706.2000.890203.x>
- Pilcher, J. R., Baillie, M. G. L., Schmidt, B., & Becker, B. (1984). A 7,272-year tree-ring chronology for Western Europe. *Nature*, 312(5990), 150–152. <https://doi.org/10.1038/312150a0>
- Pilowsky, J. A., Manica, A., Brown, S., Rahbek, C., & Fordham, D. A. (2022). Simulations of human migration into North America are more sensitive to demography than choice of palaeoclimate model. *Ecological Modelling*, 473(September), 110115. <https://doi.org/10.1016/j.ecolmodel.2022.110115>
- Plicht, J. van der, Ramsey, C. B., Heaton, T. J., Scott, E. M., & Talamo, S. (2020). Recent developments in calibration for archaeological and environmental samples. *Radiocarbon*, 62(4), 1095–1117. <https://doi.org/10.1017/RDC.2020.22>
- R Core Team. (2020). *A Language and Environment for Statistical Computing* (Vol. 3, pp. <https://www.R-project.org>). R Foundation for Statistical Computing. <http://www.r-project.org>
- Ramrath, A., Zolitschka, B., Wulf, S., & Negendank, J. F. W. (1999). Late pleistocene climatic variations as recorded in two Italian maar lakes (Lago di Mezzano, Lago Grande di Monticchio). *Quaternary Science Reviews*, 18(7), 977–992. [https://doi.org/10.1016/S0277-3791\(99\)00009-8](https://doi.org/10.1016/S0277-3791(99)00009-8)
- Ramsey, C. B. (2009a). Bayesian analysis of radiocarbon dates. *Radiocarbon*, 51(1), 337–360. <https://doi.org/10.1017/s0033822200033865>

- Ramsey, C. B. (2009b). Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon*, 51(3), 1023–1045. <https://doi.org/10.1017/s0033822200034093>
- Rasmussen, S. O., Bigler, M., Blockley, S. P., Blunier, T., Buchardt, S. L., Clausen, H. B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S. J., Fischer, H., Gkinis, V., Guillevic, M., Hoek, W. Z., Lowe, J. J., Pedro, J. B., Popp, T., Seierstad, I. K., Steffensen, J. P., Svensson, A., ... Winstrup, M. (2014). A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: Refining and extending the INTIMATE event stratigraphy. *Quaternary Science Reviews*, 106, 14–28. <http://doi.org/10.1016/j.quascirev.2014.09.007>
- Rebollo, N. R., Weiner, S., Brock, F., Meignen, L., Goldberg, P., Belfer-Cohen, A., Bar-Yosef, O., & Boaretto, E. (2011). New radiocarbon dating of the transition from the Middle to the Upper Paleolithic in Kebara Cave, Israel. *Journal of Archaeological Science*, 38(9), 2424–2433. <https://doi.org/10.1016/j.jas.2011.05.010>
- Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Hajdas, I., Heaton, T. J., Hogg, A. G., Hughen, K. A., Kromer, B., Manning, S. W., Muscheler, R., ... Talamo, S. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon*, 62(4), 725–757. <https://doi.org/10.1017/RDC.2020.41>
- Reynolds, N., & Green, C. (2019). Spatiotemporal modelling of radiocarbon dates using linear regression does not indicate a vector of demic dispersal associated with the earliest Gravettian assemblages in Europe. *Journal of Archaeological Science: Reports*, 27(March), 101958. <https://doi.org/10.1016/j.jasrep.2019.101958>
- Reynolds, N., Lisitsyn, S. N., Sablin, M. V., Barton, N., & Higham, T. (2015). Chronology of the European Russian Gravettian: New radiocarbon dating results and interpretation. *Quartar*, 62, 121–132. https://doi.org/10.7485/QU62_5
- Riel-Salvatore, J., & Negrino, F. (2018). Human adaptations to climatic change in Liguria across the Middle–Upper Paleolithic transition. *Journal of Quaternary Science*, 33(3), 313–322. <https://doi.org/10.1002/jqs.3005>
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., GuilleráArroita, G., Hauenstein, S., LahozMonfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Crossvalidation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure.

- Ecography*, 40(8), 913–929. <https://doi.org/10.1111/ecog.02881>
- Roebroeks, W., & Vill, P. (2011). On the earliest evidence for habitual use of fire in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 108(13), 5209–5214. <https://doi.org/10.1073/pnas.1018116108>
- Ronchitelli, A., Mugnaini, S., Arrighi, S., Atrei, A., Capecchi, G., Giamello, M., Longo, L., Marchettini, N., Viti, C., & Moroni, A. (2015). When technology joins symbolic behaviour: The Gravettian burials at Grotta Paglicci (Rignano Garganico - Foggia - Southern Italy). *Quaternary International*, 359, 423–441. <https://doi.org/10.1016/j.quaint.2014.08.038>
- Ruth, U., Wagenbach, D., Steffensen, J. P., & Bigler, M. (2003). Continuous record of microparticle concentration and size distribution in the central Greenland NGRIP ice core during the last glacial period. *Journal of Geophysical Research: Atmospheres*, 108(D3), n/a–n/a. <https://doi.org/10.1029/2002JD002376>
- Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., & Perera, B. J. C. (2018). Statistical downscaling of precipitation using machine learning techniques. *Atmospheric Research*, 212(May), 240–258. <https://doi.org/10.1016/j.atmosres.2018.05.022>
- Sakamoto, M., Kodaira, A., & Imamura, M. (2004). An automated AAA preparation system for AMS radiocarbon dating. *Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms*, 223–224(SPEC. ISS.), 298–301. <https://doi.org/10.1016/j.nimb.2004.04.059>
- Sakoe, H., & Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>
- Salonen, J. S., Korpela, M., Williams, J. W., & Luoto, M. (2019). Machine-learning based reconstructions of primary and secondary climate variables from North American and European fossil pollen data. *Scientific Reports*, 9(1), 15805. <https://doi.org/10.1038/s41598-019-52293-4>
- Saltr , F., Rodr guez-Rey, M., Brook, B. W., Johnson, C. N., Turney, C., Alroy, J., Cooper, A., Beeton, N., Bird, M. I., Fordham, D. A., Gillespie, R., Herrando-P rez, S., Jacobs, Z., Miller, G. H., Nogu s-Bravo, D., Prideaux, G. J., Roberts, R. G., & Bradshaw, C. J. A. (2016). Climate change not to blame for late Quaternary megafauna extinctions in Australia. *Nature Communications*, 7(1), 10511. <https://doi.org/10.1038/ncomms10511>
- Sanchez Go ni, M. F., & Harrison, S. P. (2010). Millennial-scale climate variability and vegetation changes during the Last Glacial: Concepts and terminology. *Quaternary Science Reviews*, 29(21–22), 2823–2827. <https://doi.org/10.1>

016/j.quascirev.2009.11.014

- Sánchez Goñi, M. F. (2022). *The climatic and environmental context of the Late Pleistocene* (Vol. 5, pp. 17–38). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-821428-2.00017-2>
- Sánchez Goñi, M. F., Desprat, S., Daniau, A.-L., Bassinot, F. C., Polanco-Martínez, J. M., Harrison, S. P., Allen, J. R. M., Anderson, R. S., Behling, H., Bonnefille, R., Burjachs, F., Carrión, J. S., Cheddadi, R., Clark, J. S., Combourieu-Nebout, N., Mustaphi, Colin. J. C., Debussk, G. H., Dupont, L. M., Finch, J. M., ... Yamamoto, M. (2017). The ACER pollen and charcoal database: a global resource to document vegetation and fire response to abrupt climate changes during the last glacial period. *Earth System Science Data*, 9(2), 679–695. <https://doi.org/10.5194/essd-9-679-2017>
- Sánchez Goñi, M. F., Desprat, S., Fletcher, W. J., Morales-Molino, C., Naughton, F., Oliveira, D., Urrego, D. H., & Zorzi, C. (2018). Pollen from the deep-sea: A breakthrough in the mystery of the ice ages. *Frontiers in Plant Science*, 9(January), 1–24. <https://doi.org/10.3389/fpls.2018.00038>
- Santos, G. M., Bird, M. I., Pillans, B., Fifield, L. K., Alloway, B. V., Chappell, J., Hausladen, P. A., & Arneeth, A. (2001). Radiocarbon dating of wood using different pretreatment procedures: Application to the chronology of Rotoehu ash, New Zealand. *Radiocarbon*, 43(2 PART I), 239–248. <https://doi.org/10.1017/s0033822200038066>
- Santos, G. M., & Ormsby, K. (2013). Behavioral Variability in ABA Chemical Pretreatment Close to the 14 C Age Limit. *Radiocarbon*, 55(2), 534–544. <https://doi.org/10.1017/s0033822200057660>
- Sardá-Espinosa, A. (2019). Time-series clustering in R Using the dtwclust package. *R Journal*, 11(1), 1–45. <https://doi.org/10.32614/rj-2019-023>
- Schilt, A., Baumgartner, M., Schwander, J., Buiron, D., Capron, E., Chappellaz, J., Loulergue, L., Schüpbach, S., Spahni, R., Fischer, H., & Stocker, T. F. (2010). Atmospheric nitrous oxide during the last 140,000 years. *Earth and Planetary Science Letters*, 300(1-2), 33–43. <https://doi.org/10.1016/j.epsl.2010.09.027>
- Schmidt, I., & Zimmermann, A. (2019). Population dynamics and socio-spatial organization of the Aurignacian: Scalable quantitative demographic data for western and central Europe. *PLoS ONE*, 14(2), 1–20. <https://doi.org/10.1371/journal.pone.0211562>
- Schmidt, M. W. I., & Noack, A. G. (2000). Black carbon in soils and sediments: Analysis, distribution, implications, and current challenges. *Global Biogeochem-*

- ical Cycles*, 14(3), 777–793. <https://doi.org/10.1029/1999GB001208>
- Schweingruber, F. H. (1988). *Tree Rings* (Vol. 84, pp. 299–300). Springer Netherlands. <https://doi.org/10.1007/978-94-009-1273-1>
- Shackleton, N. J., Hall, M. A., & Vincent, E. (2000). Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography*, 15(6), 565–569. <https://doi.org/10.1029/2000PA000513>
- Singarayer, J. S., & Valdes, P. (2010). High-latitude climate sensitivity to ice-sheet forcing over the last 120kyr. *Quaternary Science Reviews*, 29(1-2), 43–55. <https://doi.org/10.1016/j.quascirev.2009.10.011>
- Sirocko, F., Martínez-García, A., Mudelsee, M., Albert, J., Britzius, S., Christl, M., Diehl, D., Diansberg, B., Friedrich, R., Fuhrmann, F., Muscheler, R., Hamann, Y., Schneider, R., Schwibus, K., & Haug, G. H. (2021). Muted multidecadal climate variability in central Europe during cold stadial periods. *Nature Geoscience*, 14(9), 651–658. <https://doi.org/10.1038/s41561-021-00786-1>
- Slon, V., Mafessoni, F., Vernot, B., Filippo, C. de, Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., Douka, K., Higham, T., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Kelso, J., Meyer, M., Prüfer, K., & Pääbo, S. (2018). The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature*, 561(7721), 113–116. <https://doi.org/10.1038/s41586-018-0455-x>
- Spötl, C., & Mangini, A. (2002). Stalagmite from the Austrian Alps reveals Dansgaard-Oeschger events during isotope stage 3: Implications for the absolute chronology of Greenland ice cores. *Earth and Planetary Science Letters*, 203(1), 507–518. [https://doi.org/10.1016/S0012-821X\(02\)00837-3](https://doi.org/10.1016/S0012-821X(02)00837-3)
- Staff, R., Hardiman, M., Bronk Ramsey, C., Adolphi, F., Hare, V. J., Koutsodendrakis, A., & Pross, J. (2019). Reconciling the Greenland ice-core and radiocarbon timescales through the Laschamp geomagnetic excursion. *Earth and Planetary Science Letters*, 520, 1–9. <https://doi.org/10.1016/j.epsl.2019.05.021>
- Staff, R., Reynard, L., Brock, F., & Ramsey, C. B. (2014). Wood Pretreatment Protocols and Measurement of Tree-Ring Standards at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon*, 56(2), 709–715. <https://doi.org/10.2458/56.17449>
- Staubwasser, M., Drguin, V., Onac, B. P., Assonov, S., Ersek, V., Hoffmann, D. L., & Veres, D. (2018). Impact of climate change on the transition of Neanderthals to modern humans in Europe. *Proceedings of the National Academy of Sciences*, 115(37), 9116–9121. <https://doi.org/10.1073/pnas.1808647115>
- Stepanchuk, V. N., & Vietrov, D. O. (2021). Paglicci 24A1 and Mira II/2: Episode

- at the transition between the Early and Middle UP. *Quaternary International*, 587-588(April 2020), 277–290. <https://doi.org/10.1016/j.quaint.2021.02.020>
- Stewart, J. R., García-Rodríguez, O., Knul, M. V., Sewell, L., Montgomery, H., Thomas, M. G., & Diekmann, Y. (2019). Palaeoecological and genetic evidence for Neanderthal power locomotion as an adaptation to a woodland environment. *Quaternary Science Reviews*, 217, 310–315. <https://doi.org/10.1016/j.quascirev.2018.12.023>
- Sutikna, T., Tocheri, M. W., Morwood, M. J., Saptomo, E. W., Jatmiko, Awe, R. D., Wasisto, S., Westaway, K. E., Aubert, M., Li, B., Zhao, J. X., Storey, M., Alloway, B. V., Morley, M. W., Meijer, H. J. M., Van Den Bergh, G. D., Grün, R., Dosseto, A., Brumm, A., ... Roberts, R. G. (2016). Revised stratigraphy and chronology for *Homo floresiensis* at Liang Bua in Indonesia. *Nature*, 532(7599), 366–369. <https://doi.org/10.1038/nature17179>
- Svoboda, J. A. (2007). The Gravettian on the Middle Danube. *Paléo*, 19, 203–220. <https://doi.org/10.4000/paleo.607>
- Tagliacozzo, A., & Gala, M. (2004). L'avifauna dei livelli 24-22 (Aurignaziano e Gravettiano antico) di Grotta Paglicci: l'Aspetto ambientale e quello economico. In *Paglicci: L'aurignanziano e il gravettiano antico* (pp. 71–89).
- Talamo, S., Peresani, M., Romandini, M., Duches, R., Jéquier, C., Nannini, N., Pastoors, A., Picin, A., Vaquero, M., Weniger, G. C., & Hublin, J. J. (2014). Detecting human presence at the border of the northeastern Italian Pre-Alps. ¹⁴C dating at Rio Secco Cave as expression of the first Gravettian and the late Mousterian in the northern Adriatic region. *PLoS ONE*, 9(4). <https://doi.org/10.1371/journal.pone.0095376>
- Taller, A., & Conard, N. J. (2022). Were the Technological Innovations of the Gravettian Triggered by Climatic Change? Insights from the Lithic Assemblages from Hohle Fels, SW Germany. *PaleoAnthropology*, 1(Special Issue), 82–108.
- Telford, R. J., & Birks, H. J. B. (2009). Evaluation of transfer functions in spatially structured environments. *Quaternary Science Reviews*, 28(13-14), 1309–1316. <https://doi.org/10.1016/j.quascirev.2008.12.020>
- Tilston, E. L., Ascough, P. L., Garnett, M. H., & Bird, M. I. (2016). Quantifying charcoal degradation and negative priming of soil organic matter with a ¹⁴C-dead tracer. *Radiocarbon*, 58(4), 905–919. <https://doi.org/10.1017/RDC.2016.45>
- Timmermann, A. (2020). Quantifying the potential causes of Neanderthal extinction: Abrupt climate change versus competition and interbreeding. *Quaternary*

- Science Reviews*, 238, 106331. <https://doi.org/10.1016/j.quascirev.2020.106331>
- Timmermann, A., & Friedrich, T. (2016). Late Pleistocene climate drivers of early human migration. *Nature*, 538(7623), 92–95. <https://doi.org/10.1038/nature19365>
- Tintner, J., Preimesberger, C., Pfeifer, C., Soldo, D., Ottner, F., Wriessnig, K., Rennhofer, H., Lichtenegger, H., Novotny, E. H., & Smidt, E. (2018). Impact of Pyrolysis Temperature on Charcoal Characteristics. *Industrial and Engineering Chemistry Research*, 57(46), 15613–15619. <https://doi.org/10.1021/acs.iecr.8b04094>
- Tomiyaama, S., Minami, M., Nakamura, T., Mimura, K., & Kagi, H. (2016). Changes of chemical structure and composition of charcoal by radiocarbon pretreatments: Decontamination byaba and abox treatments. *Radiocarbon*, 58(3), 565–581. <https://doi.org/10.1017/RDC.2016.29>
- Turney, C., Bird, M. I., Fifield, L. K., Kershaw, A. P., Cresswell, R. G., Santos, G. M., Di Tada, M. L., Hausladen, P. A., & Youping, Z. (2001). Development of a robust ¹⁴C chronology for Lynch’s crater (North Queensland, Australia) using different pretreatment strategies. *Radiocarbon*, 43(1), 45–54. <https://doi.org/10.1017/S0033822200031611>
- Turney, C., Kershaw, A. P., Moss, P., Bird, M. I., Fifield, L. K., Cresswell, R. G., Santos, G. M., Di Tada, M. L., Hausladen, P. A., & Zhou, Y. (2001). Redating the onset of burning of Lynch’s Crater (North Queensland): Implications for human settlement in Australia. *Journal of Quaternary Science*, 16(8), 767–771. <https://doi.org/10.1002/jqs.643>
- Tzedakis, P. C., Frogley, M. R., Lawson, I. T., Preece, R. C., Cacho, I., & Abreu, L. de. (2004). Ecological thresholds and patterns of millennial-scale climate variability: The response of vegetation in Greece during the last glacial period. *Geology*, 32(2), 109–112. <https://doi.org/10.1130/G20118.1>
- Tzedakis, P. C., Lawson, I. T., Frogley, M. R., Hewitt, G. M., & Preece, R. C. (2002). Buffered tree population changes in a Quaternary refugium: Evolutionary implications. *Science*, 297(5589), 2044–2047. <https://doi.org/10.1126/science.1073083>
- Valdes, P., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., Davies-Barnard, T., Day, J. J., Farnsworth, A., Gordon, C., Hopcroft, P. O., Kennedy, A. T., Lord, N. S., Lunt, D. J., Marzocchi, A., Parry, L. M., Pope, V., Roberts, W. H. G., Stone, E. J., ... Williams, J. H. T. (2017). The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geosci-*

- entific Model Development*, 10(10), 3715–3743. <https://doi.org/10.5194/gmd-10-3715-2017>
- Van Deusen, P. C. (1990). A dynamic program for cross-dating tree rings. *Canadian Journal of Forest Research*, 20(2), 200–205. <https://doi.org/10.1139/x90-028>
- Vandeputte, K., Moens, L., Dams, R., & Van Der Plicht, J. (1998). Study of the ^{14}C -contamination potential of C-impurities in CuO and Fe. *Radiocarbon*, 40(1), 103–110. <https://doi.org/10.1017/s003382220001794x>
- Varela, S., Lobo, J. M., & Hortal, J. (2011). Using species distribution models in paleobiogeography: A matter of data, predictors and concepts. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 310(3-4), 451–463. <https://doi.org/10.1016/j.palaeo.2011.07.021>
- Vermeersch, P. M. (2020). Radiocarbon Palaeolithic Europe database: A regularly updated dataset of the radiometric data regarding the Palaeolithic of Europe, Siberia included. *Data in Brief*, 31, 105793. <https://doi.org/10.1016/j.dib.2020.105793>
- Villa, P., Pollarolo, L., Conforti, J., Marra, F., Biagioni, C., Degano, I., Lucejko, J. J., Tozzi, C., Pennacchioni, M., Zanchetta, G., Nicosia, C., Martini, M., Sibilis, E., & Panzeri, L. (2019). *From Neandertals to modern humans: New data on the Uluzzian* (Vol. 13). <https://doi.org/10.1371/journal.pone.0196786>
- Voelker, A. H. L. (2002). Global distribution of centennial-scale records for Marine Isotope Stage (MIS) 3: A database. *Quaternary Science Reviews*, 21(10), 1185–1212. [https://doi.org/10.1016/S0277-3791\(01\)00139-1](https://doi.org/10.1016/S0277-3791(01)00139-1)
- Vrac, M., Marbaix, P., Paillard, D., & Naveau, P. (2007). Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe. *Climate of the Past*, 3(4), 669–682. <https://doi.org/10.5194/cp-3-669-2007>
- Vrac, M., & Naveau, P. (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research*, 43(7), 1–13. <https://doi.org/10.1029/2006WR005308>
- Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B. M., Stewart, J. R., Pääbo, S., Collins, M. J., & Hublin, J. J. (2016). Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings of the National Academy of Sciences of the United States of America*, 113(40), 11162–11167. <https://doi.org/10.1073/pnas.1605834113>

- Wenk, C. (2003). Applying an edit distance to the matching of tree ring sequences in dendrochronology. *Journal of Discrete Algorithms*, 1(5-6), 367–385. [https://doi.org/10.1016/S1570-8667\(03\)00028-5](https://doi.org/10.1016/S1570-8667(03)00028-5)
- Wierer, U. (2013). Variability and standardization: The early Gravettian lithic complex of Grotta Paglicci, Southern Italy. *Quaternary International*, 288, 215–238. <https://doi.org/10.1016/j.quaint.2012.04.043>
- Wigley, T. M. L., Briffa, K. R., & Jones, P. D. (1984). On the average value of correlated time series with applications in dendroclimatology and hydrometeorology. *Journal of Climate & Applied Meteorology*, 23(2), 201–213. [https://doi.org/10.1175/1520-0450\(1984\)023%3C0201:OTAVOC%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1984)023%3C0201:OTAVOC%3E2.0.CO;2)
- Wigley, T. M. L., Jones, P. D., & Briffa, K. R. (1987). Cross-dating methods in dendrochronology. *Journal of Archaeological Science*, 14(1), 51–64. [https://doi.org/10.1016/S0305-4403\(87\)80005-5](https://doi.org/10.1016/S0305-4403(87)80005-5)
- Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Vestergård, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L. S., Pearman, P. B., Cheddadi, R., Murray, D., Bråthen, K. A., Yoccoz, N., ... Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, 506(7486), 47–51. <https://doi.org/10.1038/nature12921>
- Willmes, C., Becker, D., Brocks, S., Hütt, C., & Bareth, G. (2017). High Resolution KöppenGeiger Classifications of Paleoclimate Simulations. *Transactions in GIS*, 21(1), 57–73. <https://doi.org/10.1111/tgis.12187>
- Wolfe, M.-N., Levavasseur, G., Daniau, A.-L., Kageyama, M., Urrego, D. H., Sánchez Goñi, M. F., & Hameed, V. (2014). Impact of precession on the climate, vegetation and fire activity in southern Africa during MIS4. *Climate of the Past*, 10(3), 1165–1182. <https://doi.org/10.5194/cp-10-1165-2014>
- Wood, R. (2015). From revolution to convention: The past, present and future of radiocarbon dating. *Journal of Archaeological Science*, 56, 61–72. <https://doi.org/10.1016/j.jas.2015.02.019>
- Wood, R., Arrizabalaga, A., Camps, M., Fallon, S., Iriarte-Chiapusso, M. J., Jones, R., Maroto, J., De la Rasilla, M., Santamaría, D., Soler, J., Soler, N., Villaluenga, A., & Higham, T. (2014). The chronology of the earliest Upper Palaeolithic in northern Iberia: New insights from L'Arbreda, Labeko Koba and La Viña. *Journal of Human Evolution*, 69(1), 91–109. <https://doi.org/10.1016/j.jhevol.2013.12.017>
- Wood, R., & Brock, F. (2011). *ORAU internal report on ABOx combustion conditions and what the procedure actually does to the charcoal sample*. Oxford

Radiocarbon Accelerator Unit.

- Wood, R., Douka, K., Boscato, P., Haesaerts, P., Sinitsyn, A., & Higham, T. (2012). Testing the ABOx-SC method: Dating known-age charcoals associated with the Campanian Ignimbrite. *Quaternary Geochronology*, 9, 16–26. <https://doi.org/10.1016/j.quageo.2012.02.003>
- Wood, S. N., Goude, Y., & Shaw, S. (2015). Generalized Additive Models for Large Data Sets. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 64(1), 139–155. <https://doi.org/10.1111/rssc.12068>
- Wulf, S., Hardiman, M., Staff, R., Koutsodendris, A., Appelt, O., Blockley, S. P., Lowe, J. J., Manning, C. J., Ottoloni, L., Schmitt, A. K., Smith, V. C., Tomlinson, E. L., Vakhrameeva, P., Knipping, M., Kotthoff, U., Milner, A. M., Müller, U. C., Christanis, K., Kalaitzidis, S., ... Pross, J. (2018). The Marine Isotope Stage 1–5 cryptotephra record of Tenaghi Philippon, Greece: Towards a detailed tephrostratigraphic framework for the Eastern Mediterranean region. *Quaternary Science Reviews*, 186, 236–262. <https://doi.org/10.1016/j.quascirev.2018.03.011>
- Wulf, S., Kraml, M., Brauer, A., Keller, J., & Negendank, J. F. W. (2004). Tephrochronology of the 100ka lacustrine sediment record of Lago Grande di Monticchio (southern Italy). *Quaternary International*, 122(1 SPEC. ISS.), 7–30. <https://doi.org/10.1016/j.quaint.2004.01.028>
- Zackrisson, O., Nilsson, M.-C., & Wardle, D. A. (1996). Key ecological function of charcoal from wildfire in the Boreal forest. *Oikos*, 77(1), 10. <https://doi.org/10.2307/3545580>
- Zhang, Q. B., & Qiu, H. (2007). A millennium-long tree-ring chronology of *Sabina przewalskii* on northeastern Qinghai-Tibetan Plateau. *Dendrochronologia*, 24(2–3), 91–95. <https://doi.org/10.1016/j.dendro.2006.10.009>
- Zheng, L., Gaire, N. P., & Shi, P. (2021). High-altitude tree growth responses to climate change across the Hindu Kush Himalaya. *Journal of Plant Ecology*, 14(5), 829–842. <https://doi.org/10.1093/jpe/rtab035>
- Zilhão, J., Anesin, D., Aubry, T., Badal, E., Cabanes, D., Kehl, M., Klasen, N., Lucena, A., Martín-Lerma, I., Martínez, S., Matias, H., Susini, D., Steier, P., Wild, E. M., Angelucci, D. E., Villaverde, V., & Zapata, J. (2017). Precise dating of the Middle-to-Upper Paleolithic transition in Murcia (Spain) supports late Neandertal persistence in Iberia. *Heliyon*, 3(11), e00435. <https://doi.org/10.1016/j.heliyon.2017.e00435>
- Zilhão, J., Banks, W. E., D’Errico, F., & Gioia, P. (2015). Analysis of site formation and assemblage integrity does not support attribution of the Uluzzian to modern

- humans at Grotta del Cavallo. *PLoS ONE*, 10(7). <https://doi.org/10.1371/journal.pone.0131181>
- Zilhão, J., D'Errico, F., Bordes, J. G., Lenoble, A., Texier, J. P., & Rigaud, J. P. (2006). Analysis of Aurignacian interstratification at the Châtelperronian- type site and implications for the behavioral modernity of Neandertals. *Proceedings of the National Academy of Sciences of the United States of America*, 103(33), 12643–12648. <https://doi.org/10.1073/pnas.0605128103>
- Zorzi, F. (1964). Palaeolithic Discoveries in the Grotta Paglicci. *Antiquity*, 38(149), 38–44. <https://doi.org/10.1017/S0003598X00068757>
-

Appendices



Dynamic Time Warping Function

The following R function applies Dynamic Time Warping to two signals (contained in the dataframes `query` and `reference`) to produce alignments shown in Chapter 3. Parameters `window.start` and `window.end` define the alignment window.

```
# Required packages: dplyr, dtw

proxydtw <- function(
  query, # Query signal: dataframe with columns depth, age, and signal
  reference, # Reference signal: dataframe with columns age and signal
  window.start = 60000, # DTW window start (year BP)
  window.end = 30000 # DTW window end (year BP)
) {

  # Save indices
  query$idx <- 1:nrow(query)
  reference$idx <- 1:nrow(reference)

  # Truncate query to window.start and window.end
  query.truncated <- query %>%
    filter(age <= window.start
           & age >= window.stop)

  # Scale reference and query signals
  x <- scale(query.truncated$signal)
  y <- scale(reference$signal)

  # Define elasticity
  elasticity <- 2 * round(length(y) / length(x)) + 1

  # Run DTW
  w <- dtw(x = x, # Query
           y = y, # Reference
           step.pattern = mvmStepPattern(elasticity), # MVM step pattern
```

```

        keep.internals = T # Store indices in DTW object
    )

    # Return results
    realigned.data <- data.frame(
        depth = query.truncated$depth,
        signal = query.truncated$signal,
        original.index = query.truncated$idx,
        original.age = query.truncated$age)

    # Add realignment
    realigned.data$realigned.age <- reference$age[w$index2]

    # Return dataframe
    return(realigned.data)
}

```

B

Correlation-based Dynamic Time Warping Functions

The following two functions were written for Chapter 4. Although the higher-level scripts and functions for this study were written in R, the core of the correlation-based DTW was outsourced to C++ to improve computational performance. The first function (`correlate`) calculates the Pearson correlation coefficient between two vectors of equal length. The second function (`correlation_dtw`) takes a query and reference vector, along with the parameters `nlp` (non-linear penalty, z in Chapter 4) and `k` (correlation window size, k), and returns the DTW distance between the two signals using a correlation-based DTW metric.

```
// Correlation function

float correlate(
    NumericVector x, // First vector
    NumericVector y // Second vector
){
    // Get length
    int n = x.length();

    // Calculate and return correlation
    NumericVector xy = x * y;
    NumericVector x_sq = x * x;
    NumericVector y_sq = y*y;

    return float (n * sum(xy) - sum(x) * sum(y)) /
        sqrt((n*sum(x_sq) - (sum(x) * sum(x))) *
            ((n * sum(y_sq)) - (sum(y) * sum(y))));
}

// DTW function
```



```

float correlation_dtw(
    NumericVector x, // Query
    NumericVector y, // Reference
    float nlp, // Non-linear penalty (z)
    float k // Correlation window size
){
    // Get lengths of query and reference
    int N = x.length();
    int M = y.length();

    // Build distance and cost matrices
    NumericMatrix distanceMatrix (N, M);
    NumericMatrix costMatrix (N - k, M - k);

    // Build indices, from 0 to k-1
    IntegerVector idx = Rcpp::Range(0, (k - 1));

    // Run correlation and add to distance matrix
    for (int i = 0; i < (N - k + 1); i++) {
        for (int j = 0; j < (M - k + 1); j++) {
            IntegerVector ii(idx + i);
            IntegerVector jj(idx + j);
            distanceMatrix(i,j) = correlate(x[ii], y[jj]);
        }
    }

    // Calculate cost matrix with penalty
    for (int i = 0; i < (N - k - 1); i++) {
        for (int j = 0; j < (M - k - 1); j++) {
            NumericVector penalty = {costMatrix(i,j),
                                     costMatrix(i,j+1) * nlp,
                                     costMatrix(i+1,j) * nlp};
            costMatrix(i+1,j+1) = distanceMatrix(i,j) + max(penalty);
        }
    }

    // Return last element in cost matrix as DTW distance
    return costMatrix(N - k - 1, M - k - 1);
}

```



OxCal Code for Full Bayesian Model for Grotta Paglicci

The following OxCal code is for the “Full” Bayesian model for the Upper Palaeolithic sequence (Aurignacian to Gravettian) at Grotta Paglicci, using all available dates from the site, as described in Chapter 8. A variable outlier probability strategy was used, with greater confidence given to newly produced dates at the ORAU.

```
Plot()
{
  Outlier_Model("General",T(5),U(0,4),"t");
  Sequence("Paglicci")
  {
    Boundary("Boundary: Aurignacian Start");
    Sequence("Sequence: Aurignacian")
    {
      Phase("Phase: 24B1")
      {
        R_Date("OxA-27841, 24B1",33030,300)
        {
          Outlier("General",0.01);
        };
        R_Date("OxA-42167, 24B1",33060,250)
        {
          Outlier("General",0.01);
        };
        R_Date("UtC-1790, 24B1",34300,900)
        {
          Outlier("General",0.5);
        };
      };
      R_Date("OxA-P33375, 24A4",32350,200)
      {
        Outlier("General",0.01);
      };
      R_Date("OxA-P33354, 24A2",29460,150)
      {
```

```

    Outlier("General",0.01);
};
Phase("Phase: 24A1")
{
    R_Date("OxA-27734, 24A1",28980,240)
    {
        Outlier("General",0.01);
    };
    R_Date("UtC-1789, 24A1",29300,500)
    {
        Outlier("General",0.5);
    };
};
R_Date("OxA-42162, 24A",28940,160)
{
    Outlier("General",0.01);
};
};
Boundary("Boundary: Early Gravettian Start");
Sequence("Sequence: Early Gravettian")
{
    R_Date("OxA-27735, 23C",28580,220)
    {
        Outlier("General",0.01);
    };
    Phase("Phase: 23B")
    {
        R_Date("OxA-42163, 23B",27920,140)
        {
            Outlier("General",0.01);
        };
        R_Date("UtC-1415, 23B",26300,400)
        {
            Outlier("General",0.5);
        };
    };
    Phase("Phase: 23A")
    {
        R_Date("OxA-27736, 23A",28190,170)
        {
            Outlier("General",0.01);
        };
        R_Date("UtC-1414, 23A",28100,400)
        {
            Outlier("General",0.5);
        };
    };
    R_Date("UtC-1413, 22F4",28300,400)
    {
        Outlier("General",0.5);
    };
    R_Date("OxA-43241, 22F2",28090,140)
    {
        Outlier("General",0.01);
    };
    Phase("Phase: 22E")
    {
        R_Date("OxA-28383, 22E",27900,180)
        {
            Outlier("General",0.2);
        };
        R_Date("OxA-42164, 22E",28060,140)
        {
            Outlier("General",0.01);
        };
    };
    R_Date("OxA-43268, 22D",28140,120)
    {
        Outlier("General",0.01);
    };
};

```

```

};
R_Date("OxA-42165, 22C2",27700,140)
{
  Outlier("General",0.01);
};
R_Date("OxA-42166, 22C",27630,140)
{
  Outlier("General",0.01);
};
Phase("Phase: 22B")
{
  R_Date("OxA-43242, 22B",27960,130)
  {
    Outlier("General",0.01);
  };
  R_Date("UtC-1412, 22B",26800,300)
  {
    Outlier("General",0.5);
  };
};
R_Date("OxA-43243, 22A",26100,130)
{
  Outlier("General",0.01);
};
};
Boundary("Boundary: Evolved Gravettian Start");
Sequence("Sequence: Evolved Gravettian")
{
  Phase("Phase: 21D")
  {
    R_Date("F-55, 21D",24720,420)
    {
      Outlier("General",0.5);
    };
    R_Date("OxA-43244, 21D",25790,120)
    {
      Outlier("General",0.01);
    };
  };
  Phase("Phase: 21C")
  {
    R_Date("F-53, 21C",24210,410)
    {
      Outlier("General",0.5);
    };
    R_Date("F-54, 21C",23750,390)
    {
      Outlier("General",0.5);
    };
    R_Date("F-56, 21C",24219,410)
    {
      Outlier("General",0.5);
    };
  };
  R_Date("F-52, 21B",23470,370)
  {
    Outlier("General",0.5);
  };
  Phase("Phase: 21A1")
  {
    R_Date("OxA-27737, 21A1",23860,130)
    {
      Outlier("General",0.2);
    };
    R_Date("OxA-P33365, 21A1",24861,77)
    {
      Outlier("General",0.01);
    };
  };
};

```

```

R_Date("F-51, 21A",23040,380)
{
  Outlier("General",0.5);
};
R_Date("F-50, 20DE",22630,390)
{
  Outlier("General",0.5);
};
R_Date("0xA-27738, 20C1",24480,150)
{
  Outlier("General",0.2);
};
Phase("Phase: 20C")
{
  R_Date("F-48, 20C",22220,360)
  {
    Outlier("General",0.5);
  };
  R_Date("F-49, 20C",22110,330)
  {
    Outlier("General",0.5);
  };
};
R_Date("F-47, 20B",21260,340)
{
  Outlier("General",0.5);
};
};
Boundary("Boundary: Final Gravettian Start");
Sequence("Sequence: Final Gravettian")
{
  R_Date("0xA-P33369, 19B3",23626,79)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-27551, 19B1",22530,150)
  {
    Outlier("General",0.2);
  };
  Phase("Phase: 19A")
  {
    R_Date("F-46, 19A",20730,290)
    {
      Outlier("General",0.5);
    };
    R_Date("0xA-P33370, 19A",22647,63)
    {
      Outlier("General",0.01);
    };
  };
  Phase("Phase: 18B3")
  {
    R_Date("F-45, 18B3",20160,310)
    {
      Outlier("General",0.5);
    };
    R_Date("0xA-22521, 18B3",21990,140)
    {
      Outlier("General",0.2);
    };
    R_Date("0xA-22522, 18B3",21890,140)
    {
      Outlier("General",0.2);
    };
    R_Date("0xA-43286, 18B3",21946,63)
    {
      Outlier("General",0.01);
    };
  };
};
};

```

```

Phase("Phase: 18B2")
{
  R_Date("F-43, 18B2",14260,280)
  {
    Outlier("General",0.5);
  };
  R_Date("F-44, 18B2",20200,305)
  {
    Outlier("General",0.5);
  };
  R_Date("OxA-27739, 18B2",21080,100)
  {
    Outlier("General",0.2);
  };
  R_Date("R-1324, 18B2",17200,150)
  {
    Outlier("General",0.5);
  };
};
Boundary("Boundary: Early Epigravettian");
Sequence()
{
  Boundary("=Boundary: Aurignacian Start");
  Date("Aurignacian ");
  Boundary("=Boundary: Early Gravettian Start");
};
Sequence()
{
  Boundary("=Boundary: Early Gravettian Start");
  Date("Early Gravettian ");
  Boundary("=Boundary: Evolved Gravettian Start");
};
Sequence()
{
  Boundary("=Boundary: Evolved Gravettian Start");
  Date("Evolved Gravettian ");
  Boundary("=Boundary: Final Gravettian Start");
};
Sequence()
{
  Boundary("=Boundary: Final Gravettian Start");
  Date("Final Gravettian ");
  Boundary("=Boundary: Early Epigravettian Start");
};
};

```

D

OxCal Code for Partial Bayesian Model for Grotta Paglicci

The following OxCal code is for the “Partial” Bayesian model for the Upper Palaeolithic sequence (Aurignacian to Gravettian) at Grotta Paglicci, using only newly produced dates at the site, as described in Chapter 8. A variable outlier probability strategy was used, with greater confidence given to ABOx-SC and AOx-SC dates.

```
Plot()
{
  Outlier_Model("General",T(5),U(0,4),"t");
  Sequence("Paglicci")
  {
    Boundary("Boundary: Aurignacian Start");
    Sequence("Sequence: Aurignacian")
    {
      Phase("Phase: 24B1")
      {
        R_Date("OxA-27841, 24B1",33030,300)
        {
          Outlier("General",0.01);
        };
        R_Date("OxA-42167, 24B1",33060,250)
        {
          Outlier("General",0.01);
        };
      };
      R_Date("OxA-P33375, 24A4",32350,200)
      {
        Outlier("General",0.01);
      };
      R_Date("OxA-P33354, 24A2",29460,150)
      {
        Outlier("General",0.01);
      };
    };
  };
}
```

```

R_Date("0xA-27734, 24A1",28980,240)
{
  Outlier("General",0.01);
};
R_Date("0xA-42162, 24A",28940,160)
{
  Outlier("General",0.01);
};
};
Boundary("Boundary: Early Gravettian Start");
Sequence("Sequence: Early Gravettian")
{
  R_Date("0xA-27735, 23C",28580,220)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-42163, 23B",27920,140)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-27736, 23A",28190,170)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-43241, 22F2",28090,140)
  {
    Outlier("General",0.01);
  };
  Phase("Phase: 22E")
  {
    R_Date("0xA-28383, 22E",27900,180)
    {
      Outlier("General",0.2);
    };
    R_Date("0xA-42164, 22E",28060,140)
    {
      Outlier("General",0.01);
    };
  };
  R_Date("0xA-43268, 22D",28140,120)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-42165, 22C2",27700,140)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-42166, 22C",27630,140)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-43242, 22B",27960,130)
  {
    Outlier("General",0.01);
  };
  R_Date("0xA-43243, 22A",26100,130)
  {
    Outlier("General",0.01);
  };
};
Boundary("Boundary: Evolved Gravettian Start");
Sequence("Sequence: Evolved Gravettian")
{
  R_Date("0xA-43244, 21D",25790,120)
  {
    Outlier("General",0.01);
  };
  Phase("Phase: 21A1")
  {

```



```

R_Date("OxA-27737, 21A1",23860,130)
{
    Outlier("General",0.2);
};
R_Date("OxA-P33365, 21A1",24861,77)
{
    Outlier("General",0.01);
};
};
R_Date("OxA-27738, 20C1",24480,150)
{
    Outlier("General",0.2);
};
};
Boundary("Boundary: Final Gravettian Start");
Sequence("Sequence: Final Gravettian")
{
    R_Date("OxA-P33369, 19B3",23626,79)
    {
        Outlier("General",0.01);
    };
    R_Date("OxA-27551, 19B1",22530,150)
    {
        Outlier("General",0.2);
    };
    R_Date("OxA-P33370, 19A",22647,63)
    {
        Outlier("General",0.01);
    };
    R_Date("OxA-43286, 18B3",21946,63)
    {
        Outlier("General",0.01);
    };
    R_Date("OxA-27739, 18B2",21080,100)
    {
        Outlier("General",0.2);
    };
};
Boundary("Boundary: Early Epigravettian");
};
Sequence()
{
    Boundary("=Boundary: Aurignacian Start");
    Date("Aurignacian ");
    Boundary("=Boundary: Early Gravettian Start");
};
Sequence()
{
    Boundary("=Boundary: Early Gravettian Start");
    Date("Early Gravettian ");
    Boundary("=Boundary: Evolved Gravettian Start");
};
Sequence()
{
    Boundary("=Boundary: Evolved Gravettian Start");
    Date("Evolved Gravettian ");
    Boundary("=Boundary: Final Gravettian Start");
};
Sequence()
{
    Boundary("=Boundary: Final Gravettian Start");
    Date("Final Gravettian ");
    Boundary("=Boundary: Early Epigravettian Start");
};
};

```