

Report methodology step-by-step

Contents

| | |
|--|-----------|
| Pollen data aquisition | 2 |
| FOSSILPOL | 2 |
| Harmonisation tables | 2 |
| Data pollen assembly | 3 |
| Example record | 5 |
| Detection of past human presence | 5 |
| Detection of human events | 6 |
| Detection of human events in pollen diagrams | 6 |
| Archeological radiocarbon dates and Summed Probability Densities | 7 |
| Estimation of SPD | 8 |
| Regional age cutoff | 9 |
| Selection of the distance | 9 |
| Selecting best distance per continent | 11 |
| Pollen assemblage properties (PAP) estimation | 13 |
| Pollen richness, diversity, and evenness | 13 |
| Compositional change | 14 |
| Compositional turnover | 15 |
| Rate-of-change | 16 |
| Change-points detection and density estimates | 16 |
| Equal spacing of variables | 17 |
| Paleo Climate | 18 |
| Numerical analysis | 19 |

| | |
|--|-----------|
| Numerical analysis of hypothesis 1 | 19 |
| Hierarchical variation partitioning (hVarPart) | 19 |
| Spatial changes | 20 |
| Summary across climate zone per continent | 22 |
| Temporal changes - Continental trends | 24 |
| Figure 2 | 24 |
| Temporal variation constrained by past humans | 25 |
| Figure 3 | 26 |
| Numerical analysis of hypothesis 2 | 26 |
| Multidimensional shift | 26 |
| General trends of predictors | 28 |
| hVarPart | 29 |
| Figure 4 | 30 |

The objective of this report is to give an detailed description of individual steps of the data analysis

Pollen data acquisition

FOSSILPOL

The compilation of a pollen dataset for our analysis is performed a priori. Raw pollen datasets are carefully selected with the *R-Fossilpol* package, and the guidelines to the workflow are well described in Flantua et al. 2023 and in our website Fossilpol project. Most datasets are obtained from the Neotoma Paleoecology Database. Some additional data are from private owners in areas with data gaps, which we have limited access to use. We do not have the intellectual property rights to make these data public available. Therefore, only the derivatives of the analysis can be publicly shared. Table 1 provides a summary of the settings used in the FOSSILPOL workflow to obtain a standardised project dataset. This input data are further filter during the data processing steps in HOPE to get the final collection of a standardized dataset of high data quality we can use further in our data analyses.

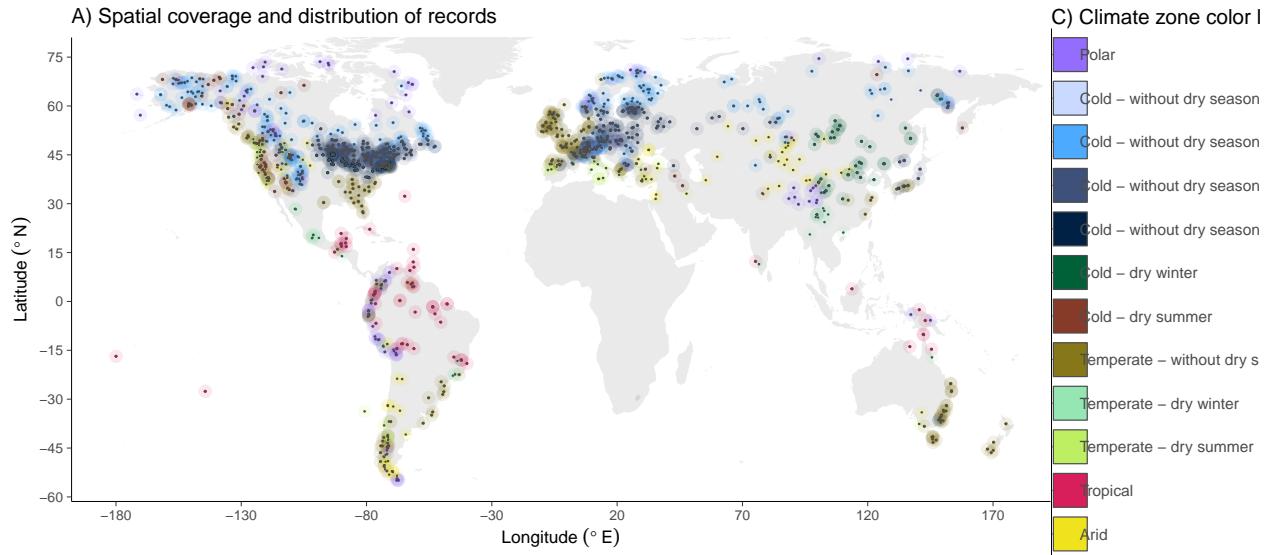
Harmonisation tables

An important step in FOSSILPOL to obtain a standardised pollen data set within and across regions is the harmonisation of pollen types. Different analysts have different backgrounds and schools using different nomenclature, and the level of pollen taxonomic identifications and names can vary widely. To be able to make numerical comparisons of different pollen records, the level of pollen taxonomy should be similar. Consequently, pollen harmonisation tables have been produced for different regions to try to minimise biases related to this. The regional harmonisation tables created in our project are for Europe, Levant, Siberia, Southern Asia, Northern America, Latin America, and the Indo-Pacific region (Birks et al. harmonisation paper). These tables are used as input in the Fossilpol workflow above (see Fossilpol step_by_step guide).

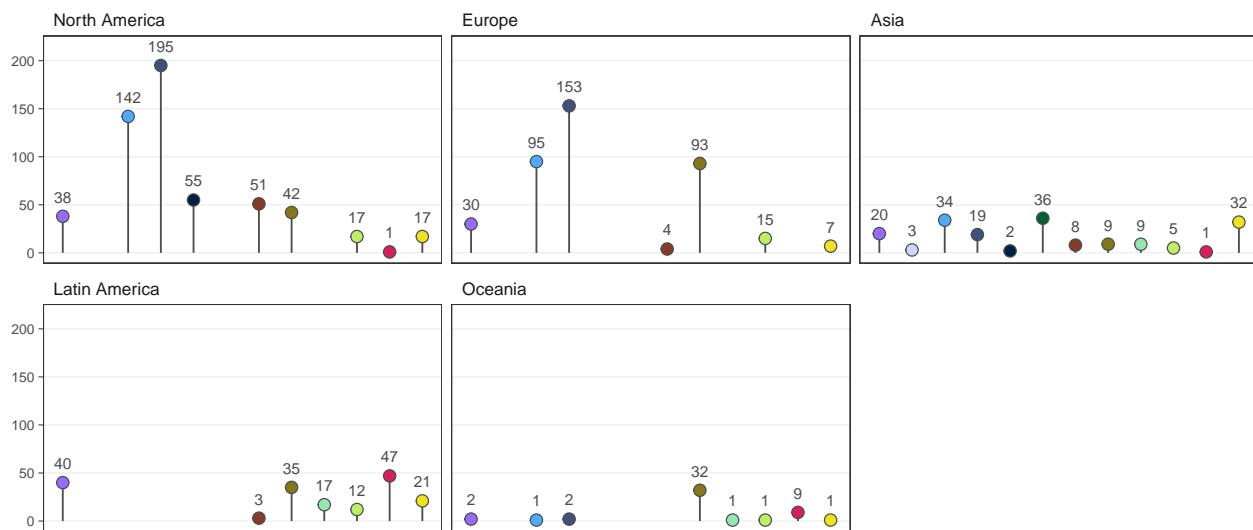
Data pollen assembly

We have applied a number of filtering criteria to obtain as high a data quality as possible so that we can compare the numerical estimates on standardised data sets. These filtering criteria are: remove potentially duplicated pollen records, sorting levels (samples) by age, remove levels (samples) lower than a threshold of total number pollen grains counted (= pollen sum), remove pollen records based on age (minimum and maximum age ranges), remove levels (samples) depending on the age of the last control point, remove samples beyond the age ranges of interest, and remove pollen records if the total number of samples (N) is too low.

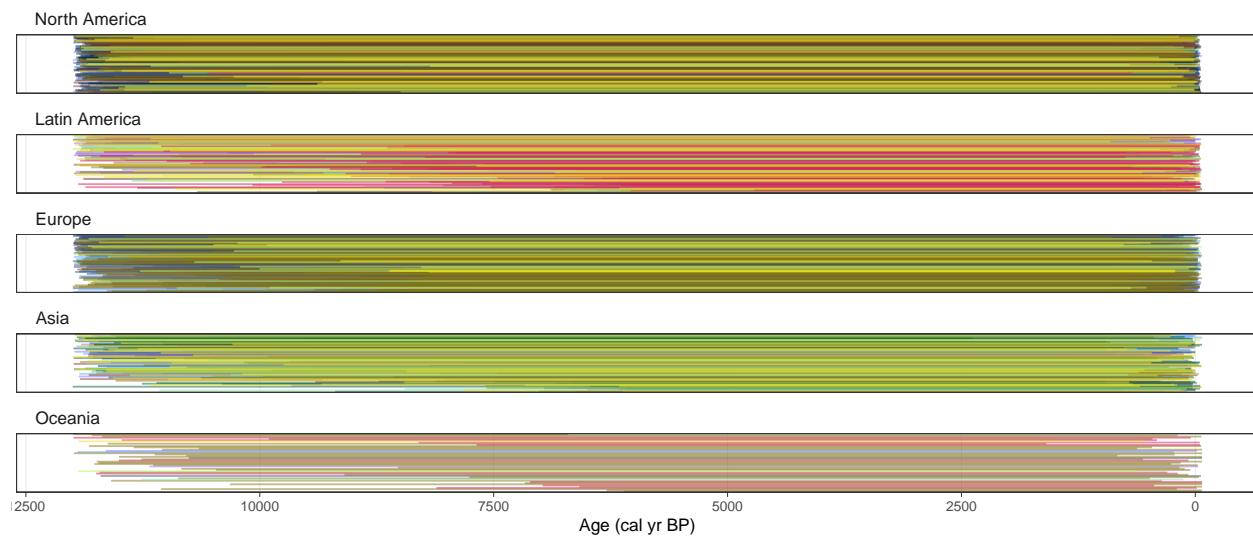
This filtering is done on the chronologies, raw pollen counts, harmonised pollen counts, and the age uncertainties from the age-depth models (Bchron). The preferable number of minimum pollen grains is set to 150, but this led to a great loss of datasets in regions with less data coverage, and we therefore reduced this number to 25 with the condition that less than 50 % of the samples must have a low pollen sum. This allow us to keep more datasets, but in the cases pollen records have a low minimum pollen sum, we acknowledge that the estimates of pollen assemblage properties (PAPs) are less robust. The maximum age beyond extrapolation is set to 3000 years because ages extrapolated beyond this threshold is considered highly uncertain. Finally, pollen records with less than 5 samples are removed for further analyses.



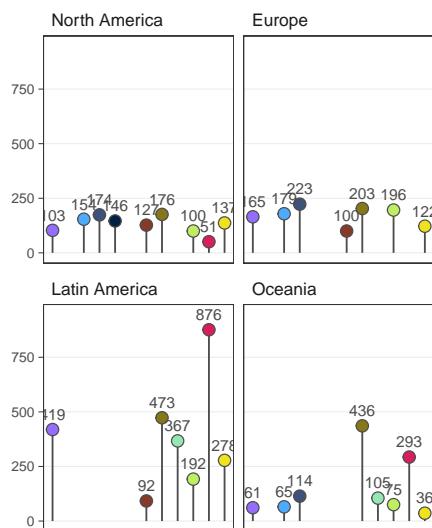
B) Number of records in each climate zone



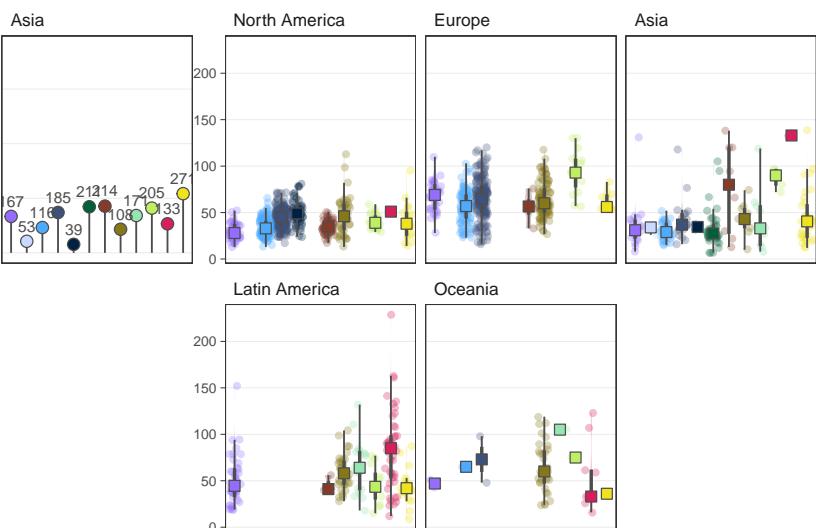
D) Temporal coverage of records



E) Total number of taxa in each climate zone

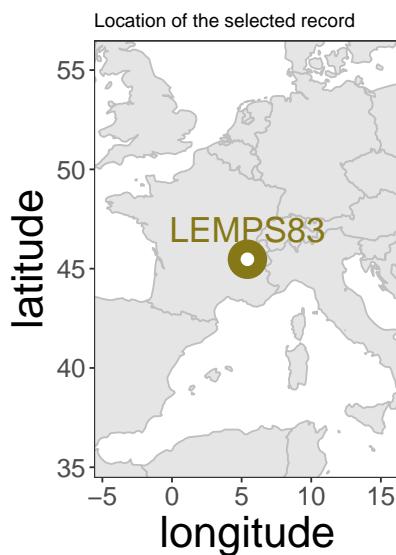


F) The number of taxa per record in each climate zone



Example record

Let's select a one record and use it as an example.



Detection of past human presence

To determine the impact of past humans on fundamental ecosystem properties, we need indicators of past human presence. We developed a method to combine the human event detection and pollen indicators of human activity identified from pollen records (based on expert knowledge) and the method for quantifying past human presence based on archaeological radiocarbon dates and Summed Probability Densities (SPD) (Bird et al. 2022). This was done to use a standardised variable as indicator of past human impact across the continents, and an attempt to reduce the potential circularity of human detection events derived from the same pollen records as the estimates of ecosystem properties.

Detection of human events

For each pollen record, we have detected periods of human presence from the pollen data. Two methods have been used:

1. detection from pollen diagrams (North America, Europe, Asia, Indopacific)
2. detection using indicator taxa (Latin America).

Detection of human events in pollen diagrams

First, a pollen diagram of each pollen record has been examined by a regional expert and the age of each event type has been recorded.

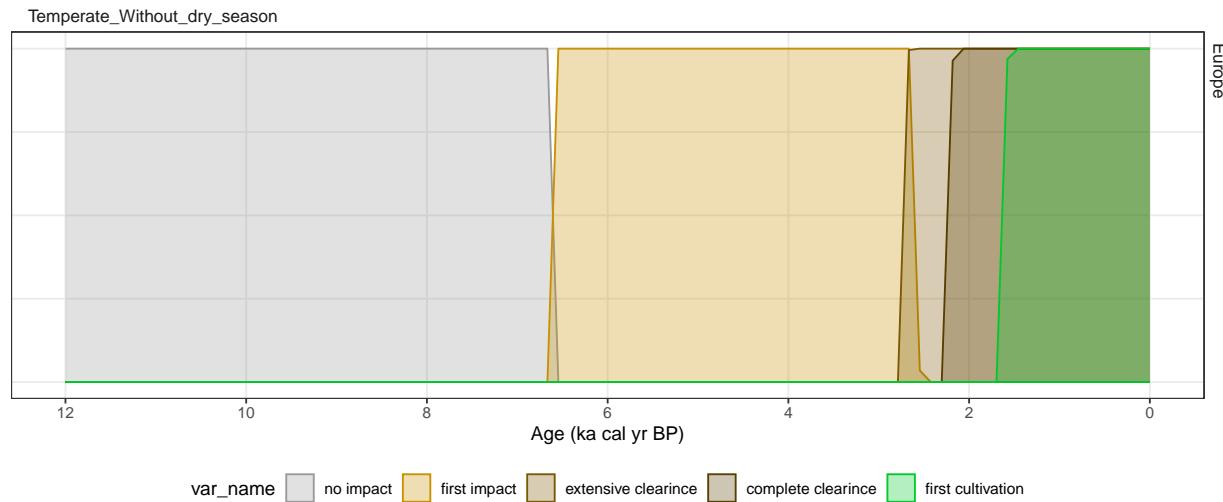
```
#>
#> -----
#>     Region           Event.type
#> -----
#>     North America   BI = Before Impact; FC = First
#>                           Cultivation; ES = European
#>                           Settlement
#>
#>     Europe          BI = Before Impact; FI = First
#>                           Indication; FCu = First
#>                           Cultivation; EC = Extensive
#>                           Clearance; CC = Complete
#>                           Clearance
#>
#>     Asia            BI = Before Impact; FI = First
#>                           Indication; FCu = First
#>                           Cultivation; EI = Extensive
#>                           Impact
#>
#>     Indopasific    no_impact = No Impact; weak =
#>                           Weak Impact; medium = Medium
#>                           Impact; strong = Strong Impact
#> -----
#>
#> Table: Table 2: Type of human events identified in pollen diagrams
```

Note that the event types are uniquely defined within continents, and human event types with the same name have different meanings between continents. More text about the human events will be added to methodology in the manuscript, see discussion notes for revision.

An algorithm is made to screen the individual pollen records in Latin America based on the information related to each pollen indicator and what human event type it is associated with. First you get binary variables (0/1) associated with each event type that is identified in each pollen record. It is complemented with a new vector with the average ages in between levels (samples) of the identified event type. This is because the time of the events is assumed to have occurred prior to the changed event. It gives a new matrix that uses the new age vector combined with the events type. The different event types are assigned binary values (0/ 1) depending on if the event type is present (i.e. when the age of the specific event is detected). Ultimately, logical principles were followed and the binary values were adjusted to get the event data for each pollen record.

Below is a figure that shows the data associated with human events in the example record. The colours represent the different types of human events in this record. The smooth trend lines represent simple binomial GAM models that show the main trends and alteration in the timing of events over time.

Temporal trend for selected record



Archeological radiocarbon dates and Summed Probability Densities

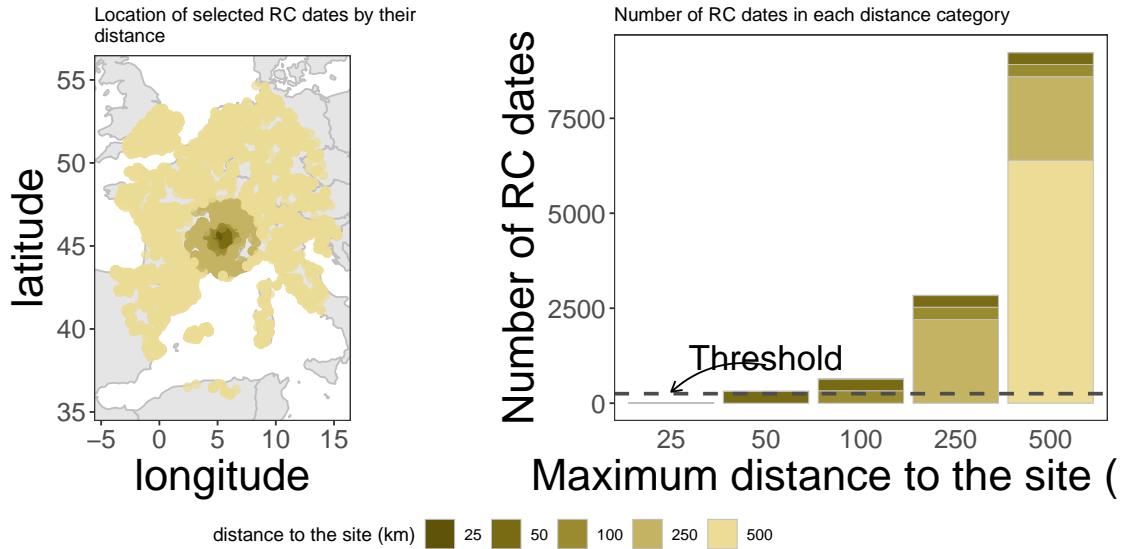
The quantification of *summed probability distribution* (SPD) requires a distance to be selected around each site location to collect the relevant archaeological radiocarbon dates around it. This will limit the area of human presence and indirectly the amount of human activity relevant to pollen records from each site.

We used the global dataset of archeological radicarbon dates (C14 dates) from Bird et al. 2022.

We have gathered the radiocarbon dates up to 500 km away and split them into groups (categories) by the distances to each of the sequence.

Only C14 dates with valid geographical location (longitude and latitude), and ‘LocAccuracy’ > 0 were selected. For each pollen record, C14 dates were classified by the geographical distance to the pollen record. The distance classes were: 5, 25, 50, 100, 250, 500 km. In each distance category, we used all C14 dates up to the maximum distance of selected category (for example: 250 contains all C14 dates up to 250 km including 5, 25,...).

However, distance class with less dates than a minimum threshold (250 RC dates) is filtered out in order to maintain robust SPD estimation.



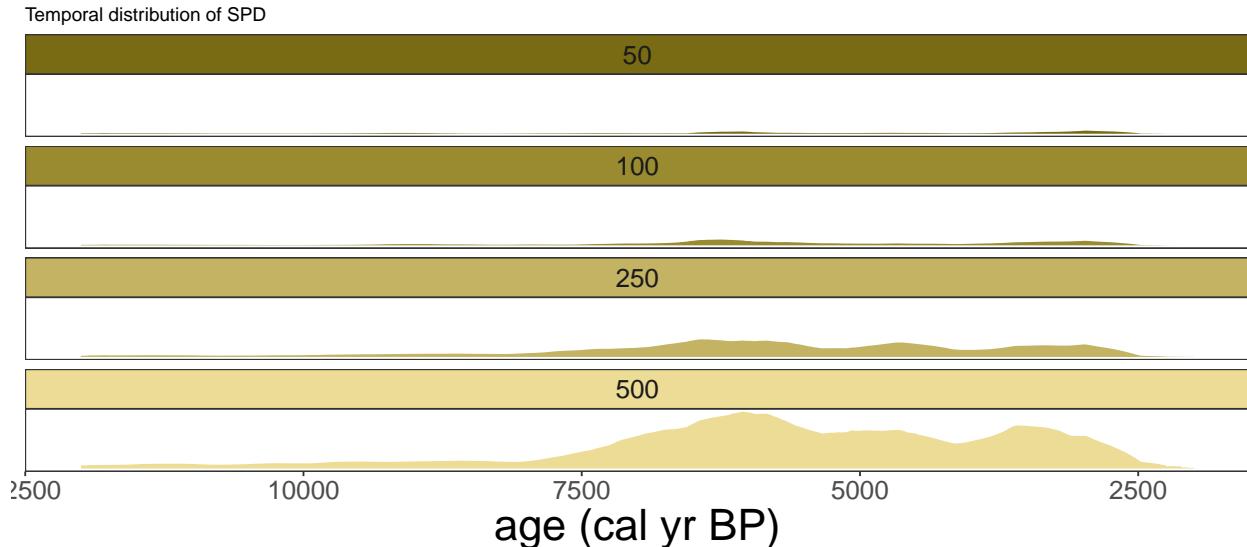
Estimation of SPD

In each distance category, SPD is estimated using `spd` function from `rcarbon` package for each year between a minimum threshold age and 12 ka.

Radiocarbon dates were calibrated using `calibrate` function from `rcarbon` package with appropriate calibration curves (“IntCal20”, “ShCal20”, “mixed”). Calibration curves were obtained `rcarbon` package and “mixed” was created using `rcarbon::mixCurves` function with ‘*p*’ = 0.5. Calibration curves were assigned by their geographical location following Hua et al., 2013.

For the selected sequences `intcal20` have been used.

The temporal distribution of SPDs constructed by different distances:



Regional age cutoff

The minimum threshold ages are different for different regions and are decided based on the availability of radiocarbon dating for the records we have in the different regions. In general, we observe there is a decrease in human presence when it increase. We considered this bias to radiocarbon dating may be limited on younger material in different subregions. Table 5 show the ages where data younger than age_from where removed.

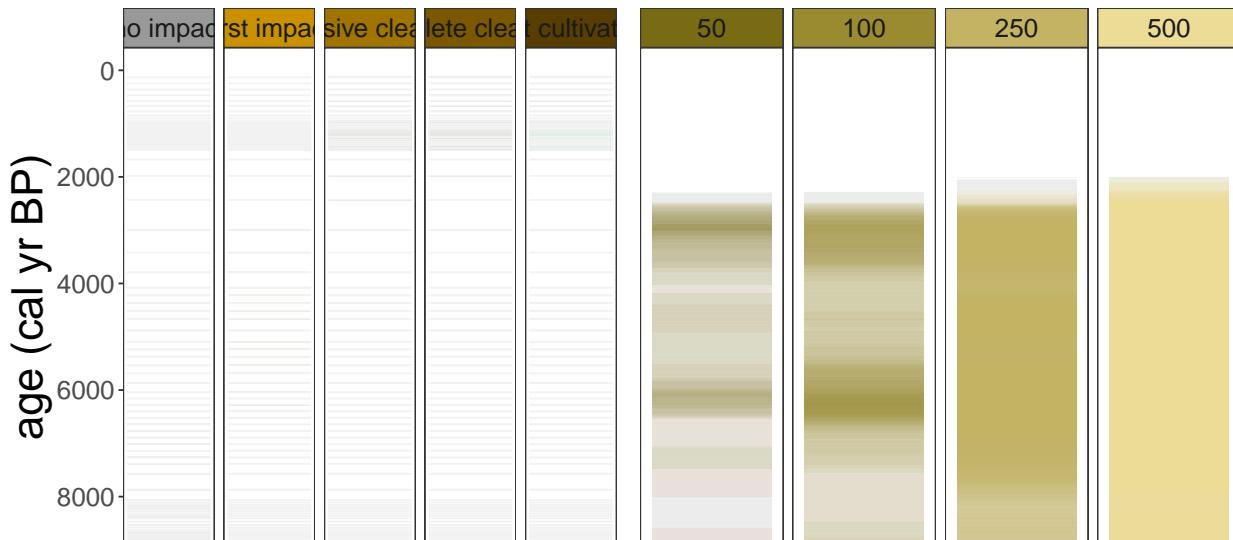
```
#>
#> -----
#>   region      age_from
#> -----
#>   Europe       2000
#>
#>   Latin America 2000
#>
#>   Asia         2000
#>
#>   North America 500
#>
#>   Oceania      500
#> -----
```

Selection of the distance

In order to select the *best distance*, we want to know, which SPD curve is the best at explaining the changes in pollen events.

Therefore, for each distance, we calculated *Redundancy Analysis* (RDA) with individual event values as response and SPD as predictors.

Here is a visualisation of the data to fit:



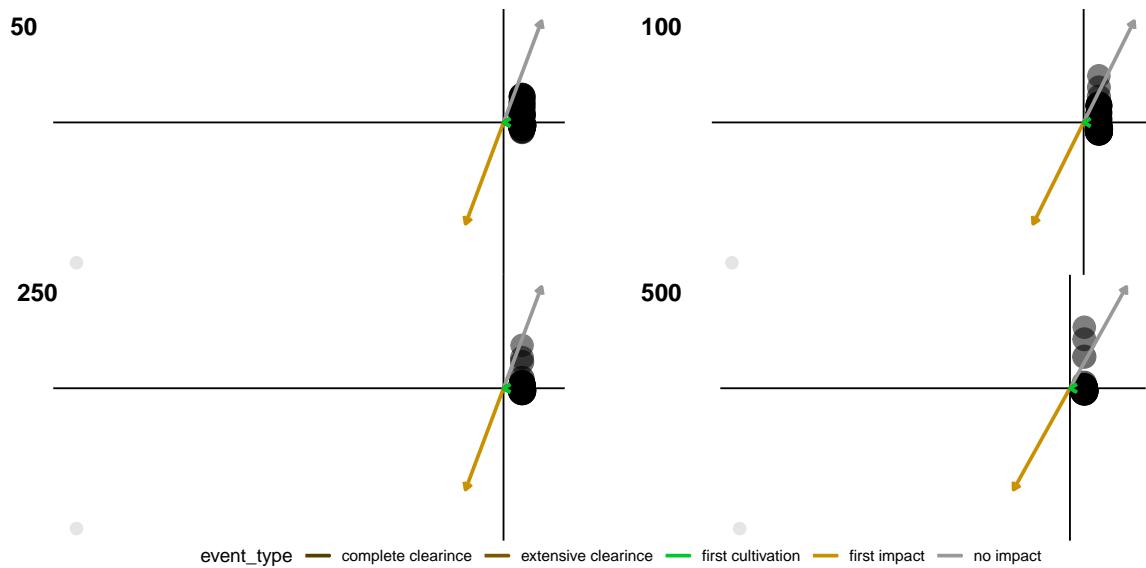
Or in a form of table:

```

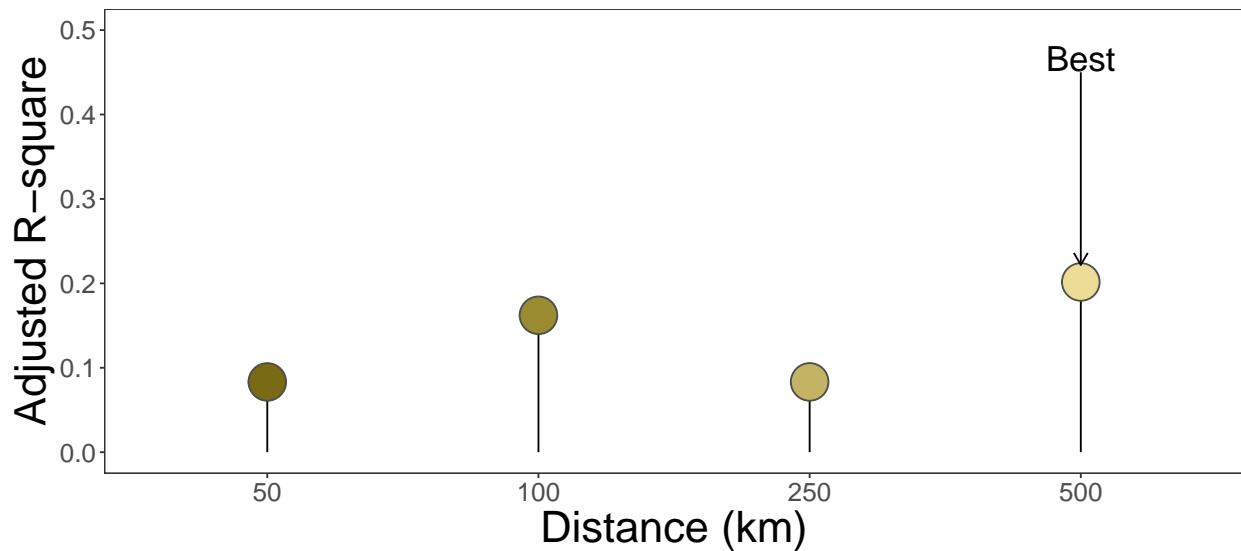
#>
#> -----
#>   age    no impact   first impact   extensive clearinace   complete clearinace
#> -----
#>   3790      0           1                   0                   0
#>
#>   6767      1           0                   0                   0
#>
#>   7015      1           0                   0                   0
#>
#>   7262      1           0                   0                   0
#>
#>   7869      1           0                   0                   0
#>
#>   8236      1           0                   0                   0
#> -----
#>
#> Table: Table continues below
#>
#>
#> -----
#>   first cultivation     50       100      250      500
#> -----
#>   0          0.02323   0.06117   0.2413   0.9766
#>
#>   0          0.01443   0.0655    0.3516   1.124
#>
#>   0          0.01545   0.05377   0.2674   0.9394
#>
#>   0          0.01738   0.04415   0.2374   0.673
#>
#>   0          0.01178   0.02182   0.1267   0.2665
#>
#>   0          0.00652   0.01761   0.08569  0.2014
#> -----

```

We can plot the individual ordination, but it is not very informative. The size and occupancy of points correspond with their ages.



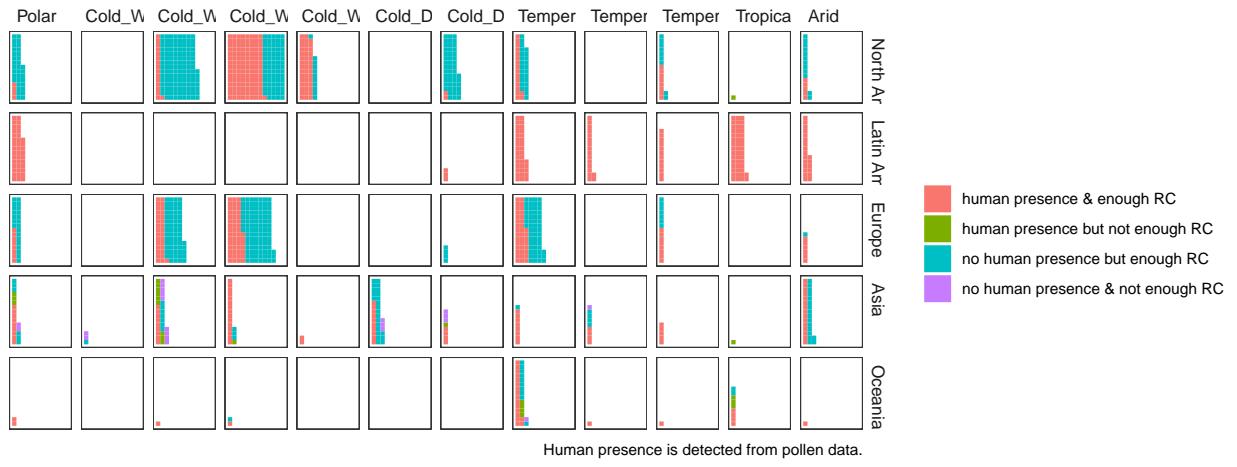
R2 We can evaluate the amount of variability explained by each ordination. We then extracted adj.r2 for each SPD distance:



Therefore we have selected **500** as the best distance and the preferred SPD curve for **4197** sequence

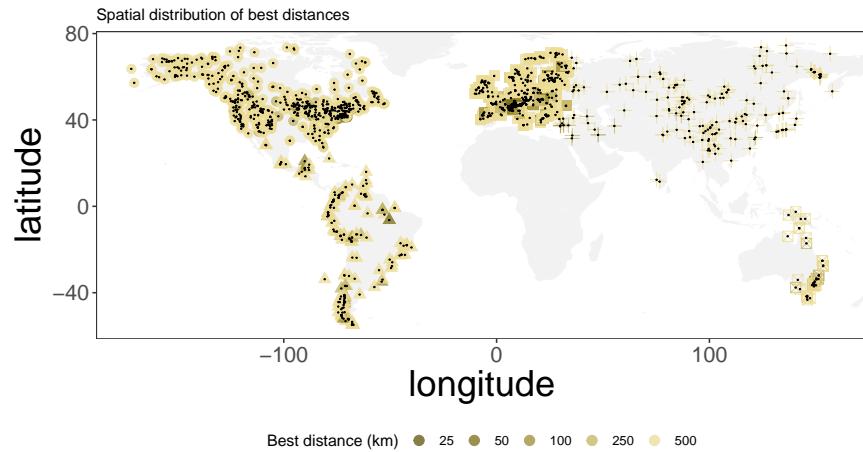
Selecting best distance per continent

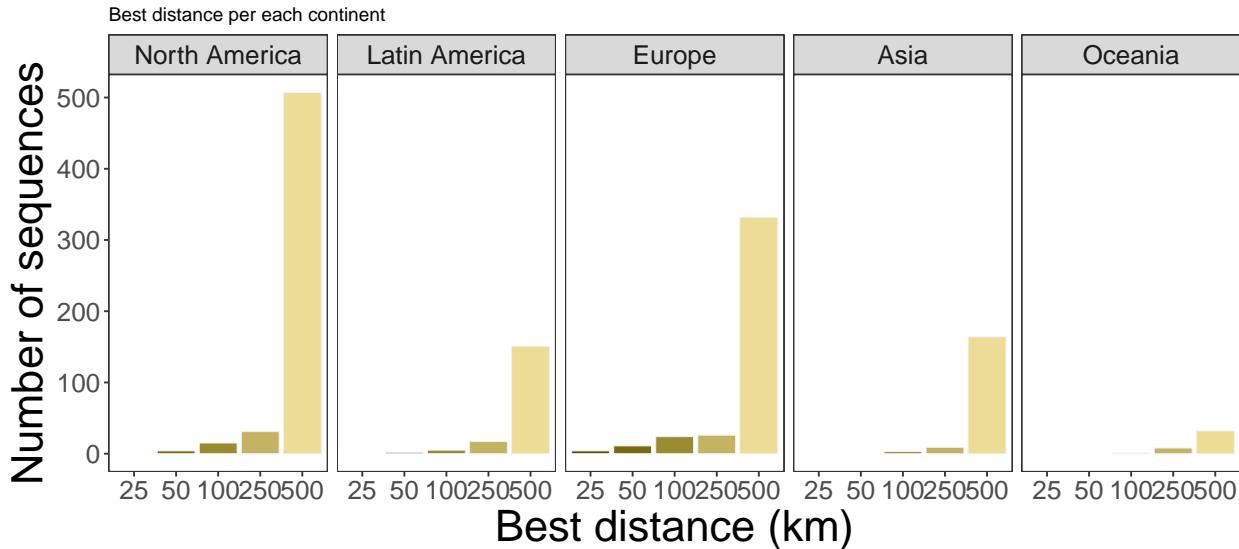
Here is a overview of the status of records base on the presence of human impact from pollen diagrams and C14 data



Now we can apply the selection of best distance to all sequences which fulfill such criteria: 1. Have events
2. Have at least one SPD constructed.

We can summary the *best distance* per each continent





Pollen assemblage properties (PAP) estimation

To prepare the response variables of our pollen dataset compilation and to be able to analyse fundamental properties of vegetation, we quantified standard estimates of pollen assemblage properties (PAP) (Bhatta et al. 2023). The PAP estimations provide different aspects of pollen assemblage diversity which includes palynological richness, diversity and evenness, compositional change and turnover, and Rate-of-Change (RoC).

These response variables are calculated using the newly developed R-Ecpol package that contain all the functions needed to estimate PAPs for our pollen data assembly. The base functions used in this package are derived from other dependency packages such as `mypart` package (Therneau et al. 2014) to estimate pollen zonations with multivariate regression trees, `vegan` (Oksanen et al. 2022) for other multivariate techniques and dissimilarity indices functions from `iNext` (Chao et al. 2014) that have been modified to extract interpolated Hill numbers based on a minimum sample size, and newly developed R functions to run DCCA using `Canoco` 4.5 (ter Braak xxxx) to list a few, among other, dependency packages. Additionally, R-Ratepol (Mottl 2021) is used get the estimates of RoC

Pollen richness, diversity, and evenness

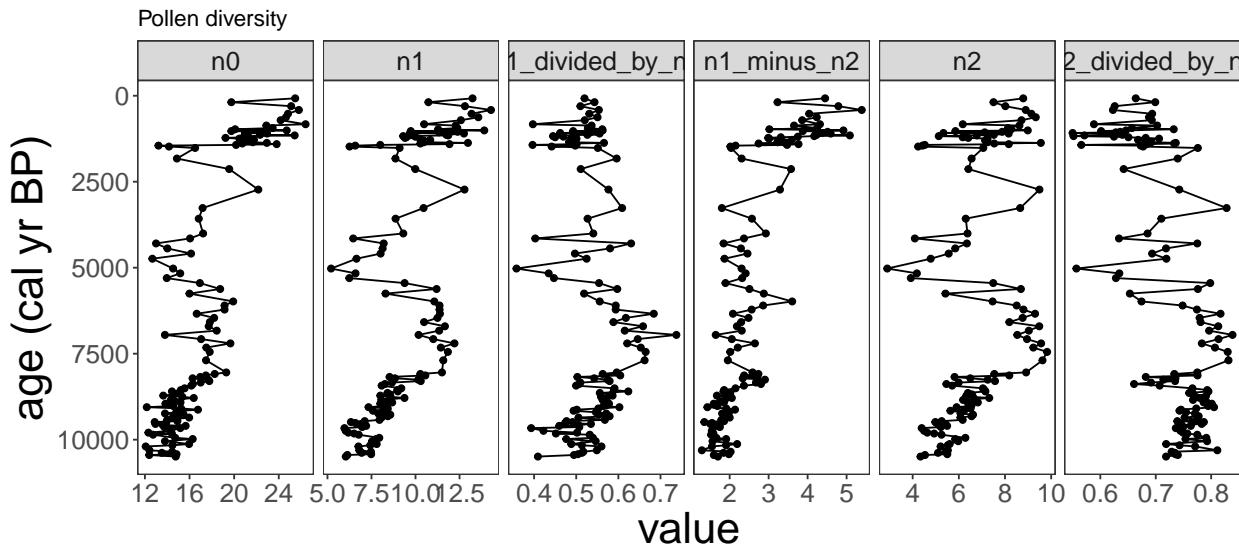
The different aspects of palynological diversity are estimated using Hill's effective species numbers N_0 , N_1 , N_2 , and the associated evenness ratios of N_2/N_1 and N_1/N_0 . These are combined through one equation where the effective species numbers differ mainly in how the rare taxa are weighted in the parameter q :

$$^qD = \left(\sum_{i=1}^S p_i^q \right)^{1/(1-q)}$$

When q is 0, rare and abundant taxa have equal weight and the number is simply the number of taxa in the sample. The equation is not possible to define for $q = 1$, but as it approaches 1, it is equal to the exponential of the well-known Shannon index and reports the number of equally common taxa. When $q = 2$, it is the same as the inverse Simpson diversity index and provides the number of equally abundant taxa with a low weight on rare taxa. The advantage of using effective species numbers is that they provide interpretable units and contain the doubling effect. To standardize the sample sizes, we use the rarefaction approach developed by Chao et al. These estimates are rarefied to the number of $n = 150$ grains, or in some cases to a lower sum (minimum $n = 25$). Some pollen records were only available as pollen percentages, and as the sample

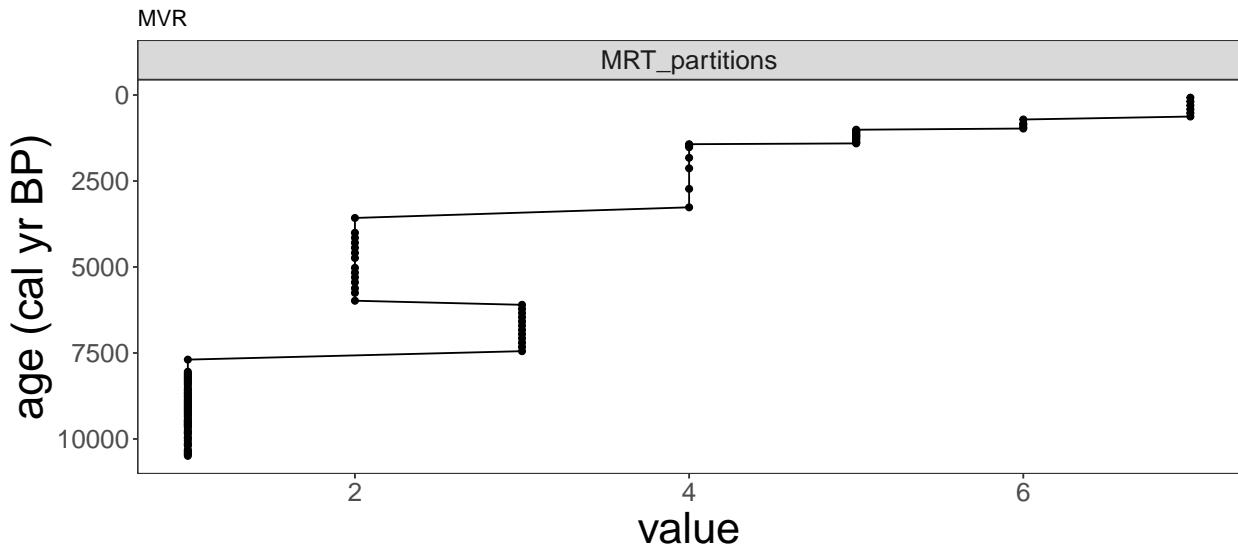
size is unknown, these are then rarefied to the minimum sum of percentages. The evenness ratios will be 1 if all taxa are equally abundant, and the ratios hence indicate changes in abundances between the numbers of rare, equally common, and abundant taxa.

We acknowledge that even though attempts are made to standardise richness and diversity estimates based on standard sample size, there are additional biases that are not taken into consideration such as differences in total pollen production and pollen representation (Odgaard 1998, 2001). In some cases, the total pollen sum is also too low to be considered a robust estimate, but it was a choice made on balancing losing too much information from geographical areas with less data coverage (see data filtering above).



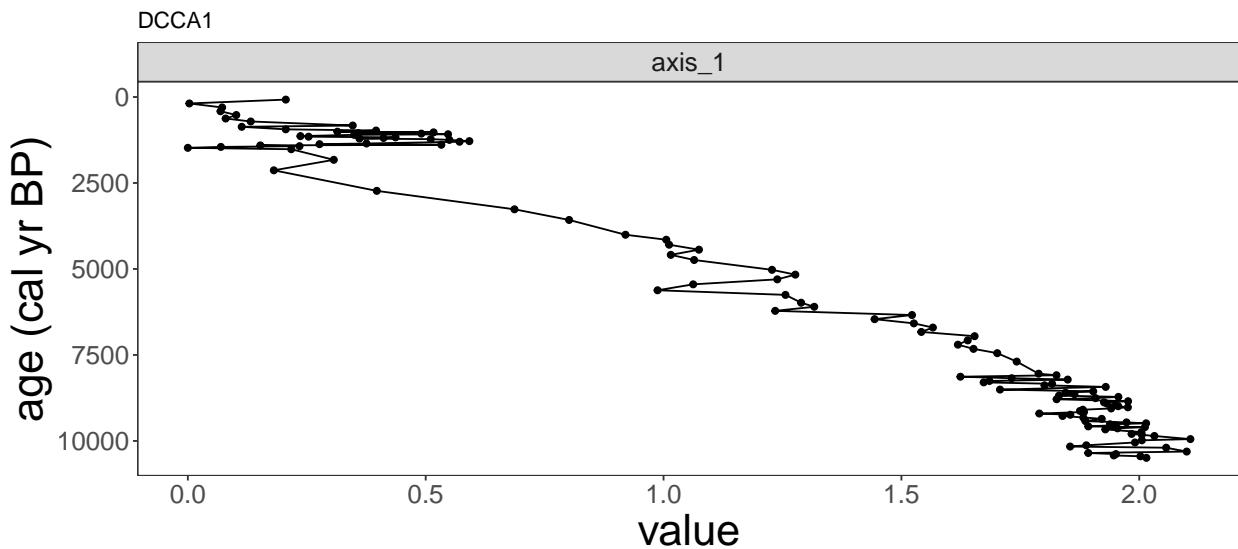
Compositional change

Compositional change is calculated using multivariate regression trees (MRT) with age as the constraining variable. MRT is in general a robust tool to explore and predict changes in multivariate data sets using environmental predictor variables (De'ath, Simpson and Birks 2012). This technique has been adopted in palaeoecology to derive zonations in pollen diagrams or shifts between periods of homogeneous vegetation in time (Simpson and Birks 2012). We use the pollen taxa in percentages without any data transformations as the response and the median ages derived from the age-depth model as the constraining variable. The recursive partitioning are based on chi-square distances between pollen samples constrained by time. The number of cross-validation is set to 1000, and the optimal sized tree is chosen based on the 1SD rule (Simpson and Birks 2012).



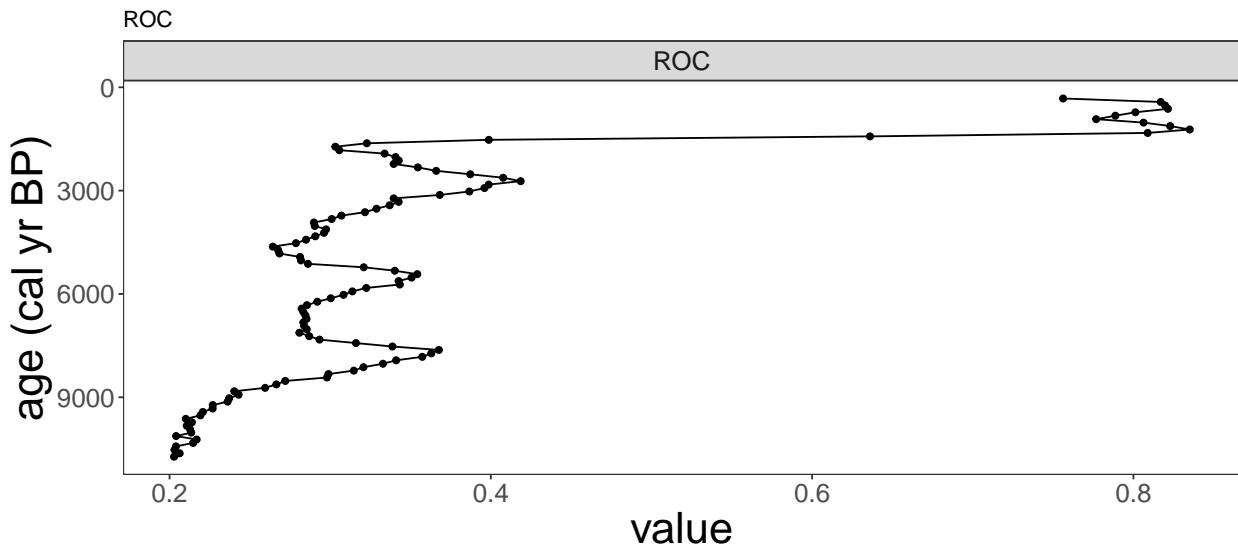
Compositional turnover

Compositional turnover is estimated using detrended canonical correspondence analysis (DCCA) with age as the explanatory variable. Changes in Weighted average (WA) sample scores (CaseR scores sensu ter Braak and Smilauer 2012) are measures of compositional turnover in standard deviation (SD) units (Birks 2007). The WA scores are regressed with time using a second-order polynomial ($\text{age} + \text{age}^2$) to allow more flexibility in the turnover pattern within a pollen record. Total compositional turnover is a measure of the total length of CaseR scores along the DCCA axis 1, whereas the pattern within a record is the measures between the individual samples along the DCCA axis 1. The response data are pollen percentages without any transformation to maintain the chi-square distances between samples, whereas the ages are the median ages derived from the age-depth model for each site.



Rate-of-change

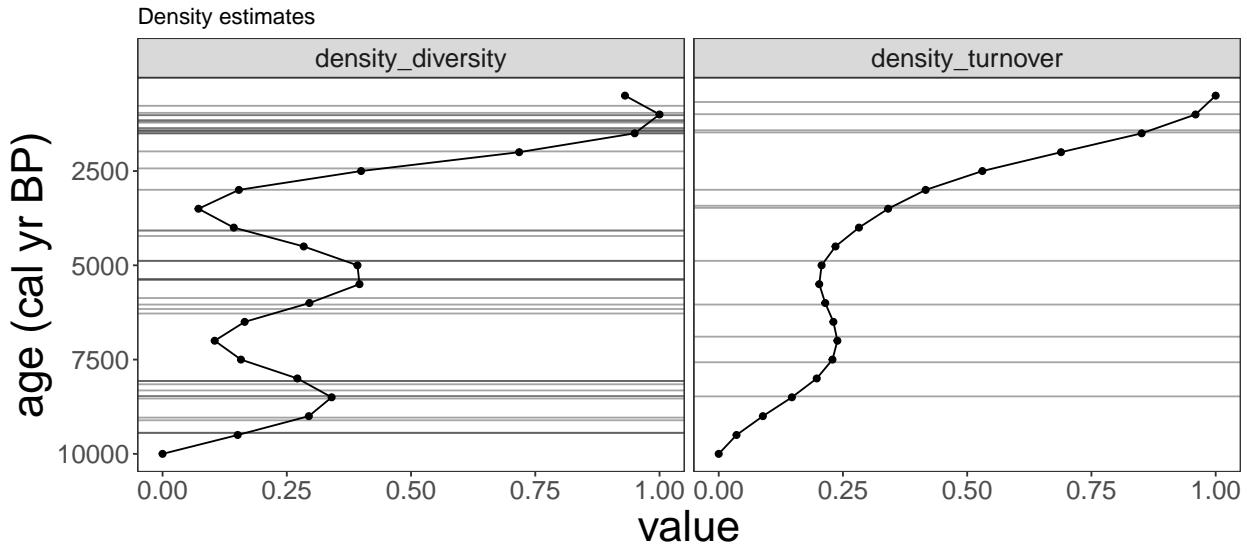
Rate-of-change for the pollen assemblages in the pollen records are quantified using the novel R-Ratepol package (Mottl et al. 2021). RoC is estimated using moving windows of 500 years' time bins of five number of windows shifts where samples are randomly selected for each bin. This approach is shown to increase the correct detection of RoC peak-points than the more traditional approaches (Mottl et al. 2021). RoC are reported as dissimilarity per 500 years using the Chord dissimilarity coefficient. Sample size is standardized in each working unit to 150 grains or the lowest number detected in each dataset. We use only the RoC scores further in the analyses.



Change-points detection and density estimates

Change-points detection of the PAP variables are calculated using conventional regression trees (RT) for single variables with Euclidean distances. The transitions between the resulting groups (or zones) per variable is detected and saved as binary (0/1) variables. A change-point is defined as 1, where the mean ages between the two consecutive samples are used as the timing of this significant change. This is done individually for each PAP variable.

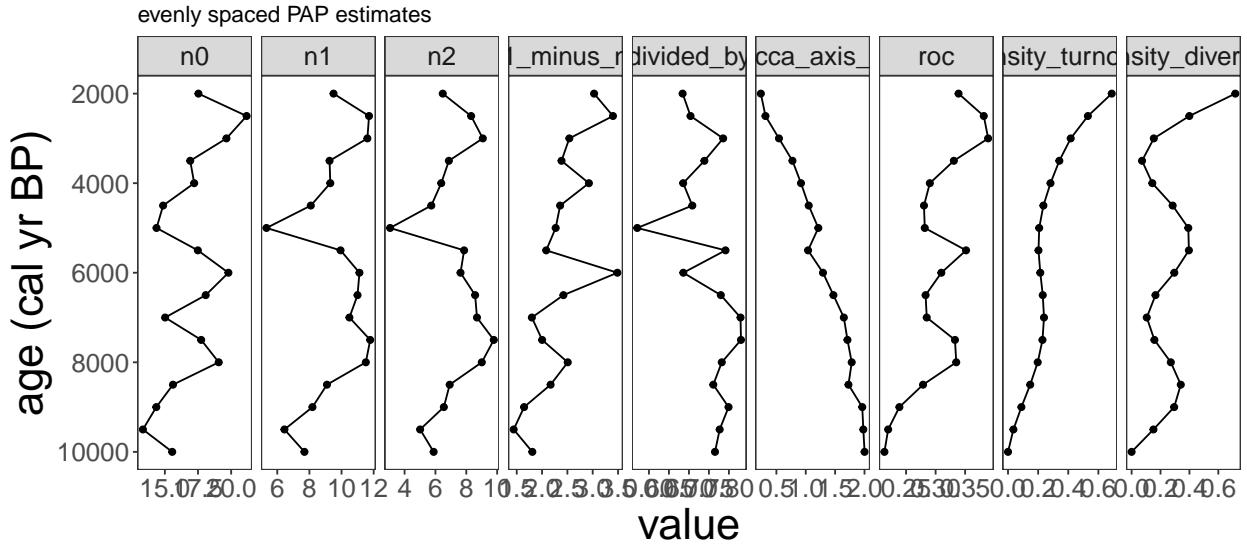
The significant change points of the richness, diversity, and evenness variables are combined into one variable, and the significant change points of compositional turnover, compositional change, and rate-of-change is combined in a second variable. The density of these two variables are calculated using a Gaussian kernel, and re-scaled to each of specific age ranges for each individual pollen record (i.e. minimum and maximum ages). To solve the boundary issue in density estimation the data is reflected to 0. We extract the interpolated values at every 500 years time step.



Equal spacing of variables

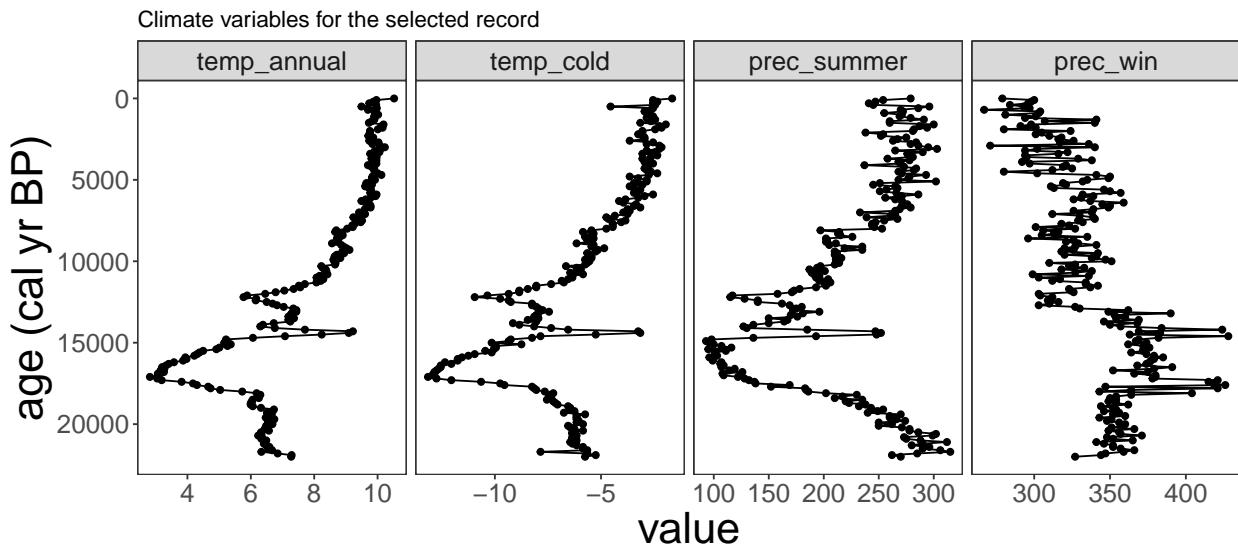
All response variables have been estimated using the harmonised pollen records for each location. To obtain estimates of equal spacing of 500 years, we used linear interpolation. In the context of the temporal analysis we analyse samples distributed in space across time, and equal time steps are necessary.

In order to choose a method of interpolation to obtain data on equal time steps, we compared generalise additive models (GAM), hierarchical generalised additive models (HGAM), and simple linear interpolation. By applying linear interpolation, we found that the correlation structure between the multivariate response variables are more similar to the original estimates without equal spacing than applying a GAM or HGAM. The GAM or HGAM models sometime showed unexpected patterns in single PAP estimate that changed these correlations. Since we cannot individually assess all the single models for each of the variable in all of the records (>1000), we choose the simplest linear interpolation method. Similar issues were detected when estimating the density variables of changes points. The first approach was to estimate densities of the individual change points, and then use hierarchical generalised additive models (HGAM) to find the common pattern between the two groups representing significant changes in richness, diversity, and evenness, or change in pollen assemblages (MRT, RoC, DCCA1). As some of the models did not converge and showed inconsistent patterns we use the density estimates on the combined variables directly and extracted the interpolated values at every 500 years step.



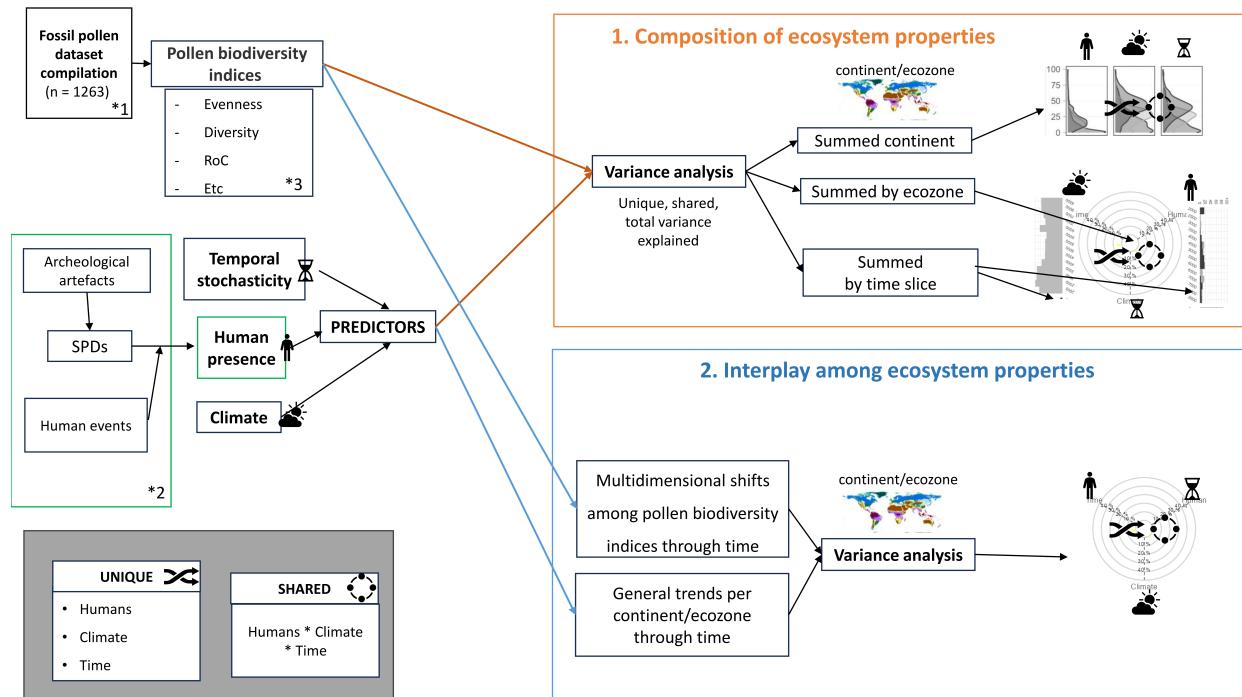
Paleo Climate

Paleoclimate from the CHELSA-TraCE21k downscaling algorithm is downloaded from the CHELSA database (Karger et al. 2021, Karger et al. 2021). The selected bioclimatic variables are annual mean temperatures °C (bio1, `temp_annual`), minimum temperatures of coldest month °C (bio6, `temp_cold`), precipitation of warmest quarter kg m⁻² quarter-1 (bio18, `prec_summer`) and precipitation of coldest quarter kg m⁻² quarter-1 (bio19, `prec_winter`), where we extracted climate values for the coordinates for each dataset_id retrieving the full time series of every 100 years. In addition, we downloaded the monthly climatology for daily maximum near-surface temperature K/10 (tasmin).



Numerical analysis

The whole project is relative complex, therefore we included a conceptual workflow figure (Figure 1). Note: Figure is work in progress and will be improved with graphic designer.



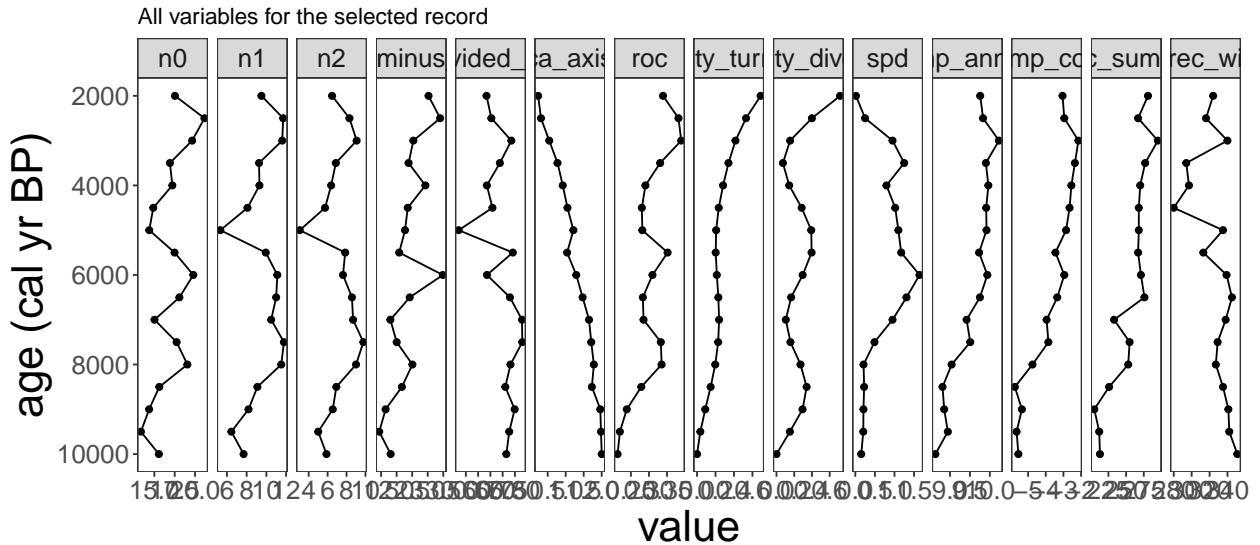
Numerical analysis of hypothesis 1

Hierarchical variation partitioning (hVarPart)

To find out if past human presence have an impact on the pollen assemblage properties, we use reduced rank multivariate regression, also known as distance-based redundancy analysis (db-RDA). We used the R package `rdacca.hp` to run hierarchical variation partitioning with several predictors. This is based on regular variation partitioning, but it estimate potentially a less biased individual variable importance (unique + shared) using the different combination of the order of predictor variables to get the average unique and shared variances in calculating the individual importance variable. We run db-RDA using *Gower-distances* as the pollen assemblage data is a mixture of different units.

Depending on the type of spatial or temporal analysis for hypothesis 1, the predictor variables are either past human impact (SPDs), the palaeoclimatic variables, and/or time (see below).

The impact of past humans (SPD) is of main interest. However, to get an understanding of the amount of human impact and the extent, it is a natural choice to compare it to climate. The palaeoclimate is represented by four variables: summer precipitation, winter precipitation, annual temperatures, and winter temperatures. These are selected as we considered them most relevant to represent major differences in climatic conditions globally (to reflect differences in warm, cold, dry, wet, or regions with high seasonality). Time is represented by the ages of each pollen record, however, this is more difficult to interpret. We assume age may represent time dependent changes such as natural successions and/or ecological changes due to interaction between taxa.



In the hierarchical variation partitioning analysis, the predictor variables can be applied either as individual predictors or as groups of predictors. In our case, we run the analysis with *groups of predictors*. This means that the palaeoclimatic variables are included as one matrix and not assessed as individual predictors. (The overall results is not very different from using individual predictors).

The analysis is run in two different ways:

1. to analyse *spatial changes* which run the hierarchical variation partitioning for single record, then summarise per continent and climate zones
2. to analyse the *temporal patterns* in space for each region of the 500 year time steps.

Spatial changes

We use the total adjusted r² to assess the goodness-of-fit of models. Adjusted r² is the modified version of r² that corrects for the number of samples and predictors using the Ezekiel formula adjr² = 1 - (1- r²)*(n-1)/(n-m-1), where n = number of samples in the dataset, and m is the number of variables. In vegan, these are so called semipartial r² (Legendre et al. 2011).

In the results, the negative adjusted r² were replaced by 0. The assessment of a satisfactory model fit is contingent upon the underlying data characteristics. Thus, we considered the range of total adjusted r² and removed models with a total fit falling below the 5% lower quantile threshold (i.e. total adj r² below 0.121).

We extracted the total variation of that whole model and the variances explained by individual predictors. There are three types of variation partition per each predictor:

- total predictor variation (= individual importance)
- shared predictor variation (= average shared partition)
- unique predictor variation (= average unique partition)

We use the total model variation to calculate the percentage of shared and unique variation explained by each predictor.

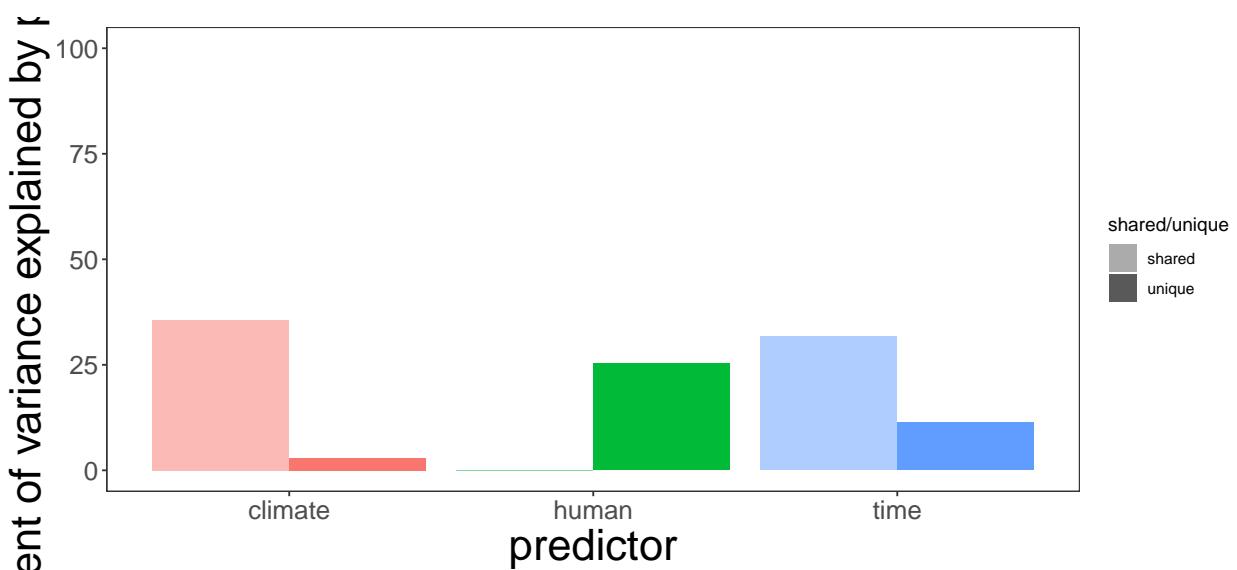
```
#>
```

```
#> -----
```

```

#>   total_variance   predictor   predictor_unique   predictor_shared
#> -----
#>   0.283           human       0.0717            0
#>
#>   0.283           climate      0.0083            0.1008
#>
#>   0.283           time        0.0321            0.0897
#> -----
#>
#> Table: Table: Result table of a single record for spatial hierachical variation partitioning (cont)
#>
#>
#> -----
#>   predictor_total   predictor_unique_percent   predictor_shared_percent
#> -----
#>   0.0521              25.34                  0
#>
#>   0.1091              2.933                 35.62
#>
#>   0.1218              11.34                 31.7
#> -----
#>
#> Table: Table continues below
#>
#>
#> -----
#>   predictor_total_percent
#> -----
#>   18.41
#>
#>   38.55
#>
#>   43.04
#> -----

```



However, for certain records, the shared variance of predictors is much larger than the total variance of the model. This caused percentages to be very high.

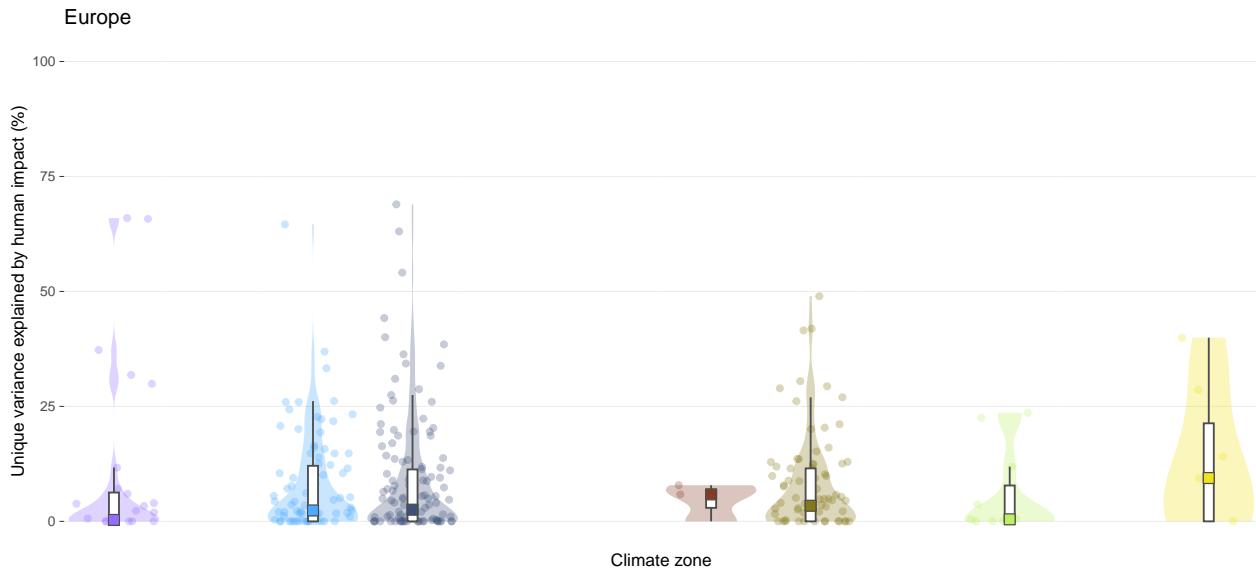
```
#>
#> -----
#>   total_variance   predictor   predictor_unique   predictor_shared
#> -----
#>   0.035           human        0                 0.0523
#>
#>   0.035           climate       0                 0.0943
#>
#>   0.035           time         0.0162            0.1104
#> -----
#>
#> Table: Table: Result table of a single record for spatial hierachical variation partitioning (cont)
#>
#>
#> -----
#>   predictor_total   predictor_unique_percent   predictor_shared_percent
#> -----
#>   0                  0                         149.4
#>
#>   0                  0                         269.4
#>
#>   0.1266             46.29                   315.4
#> -----
#>
#> Table: Table continues below
#>
#>
#> -----
#>   predictor_total_percent
#> -----
#>   0
#>
#>   0
#>
#>   361.7
#> -----
```

In these cases, the outliers above 100% were removed.

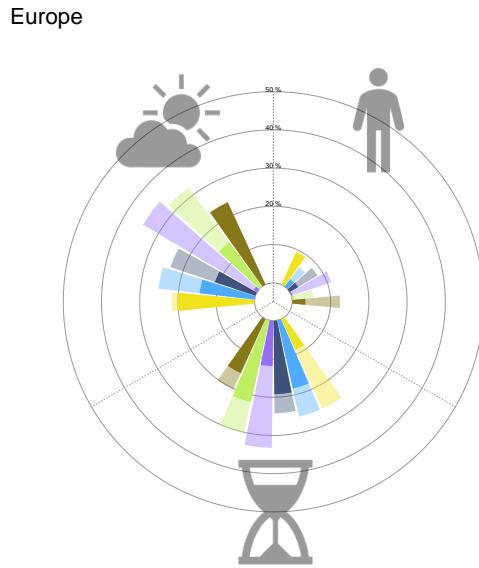
Summary across climate zone per continent

We then summarised the results per each climate zone within a continent.

Here is an example of the distribution of unique variance explained by humans for Europe (part of Figure 3).

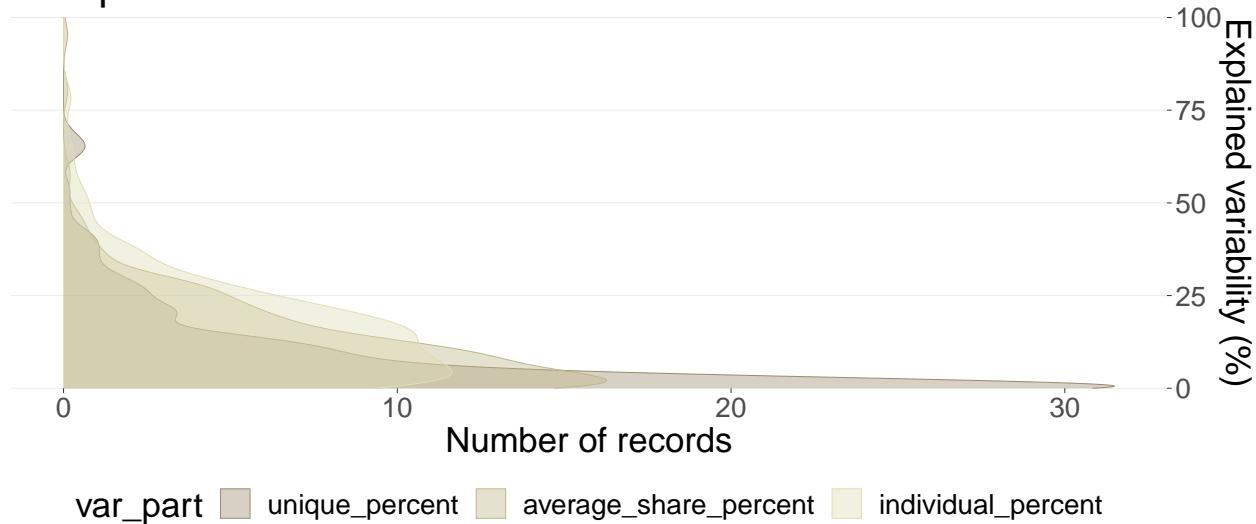


We then estimate the the median value of variance explained by each predictor (unique and shared).



In addition, we plot the distribution of other parts of variance for each predictor on the whole continent. This is part of side plots for each continents in Figure 2.

Europe – climate



Temporal changes - Continental trends

We make a new Hierarchical Variance Partition analysis, however, this time we restructure the data, so that one model is run for all response and predictors within each 500-year time slice for each continent. This means that data from all records within continent for specific time slice are placed in the same model. The predictor groups in model are the past human presence and the matrix of palaeoclimatic variables (time is not included as all sample have same time).

Similar to spatial changes, unique and shared variance explained by each predictor is extracted and plotted. In the figure, each bar (column) is a single model. The light grey bars are period, where we do not have data or model fail. The green color represent unique/shared variance explained by human predictor.

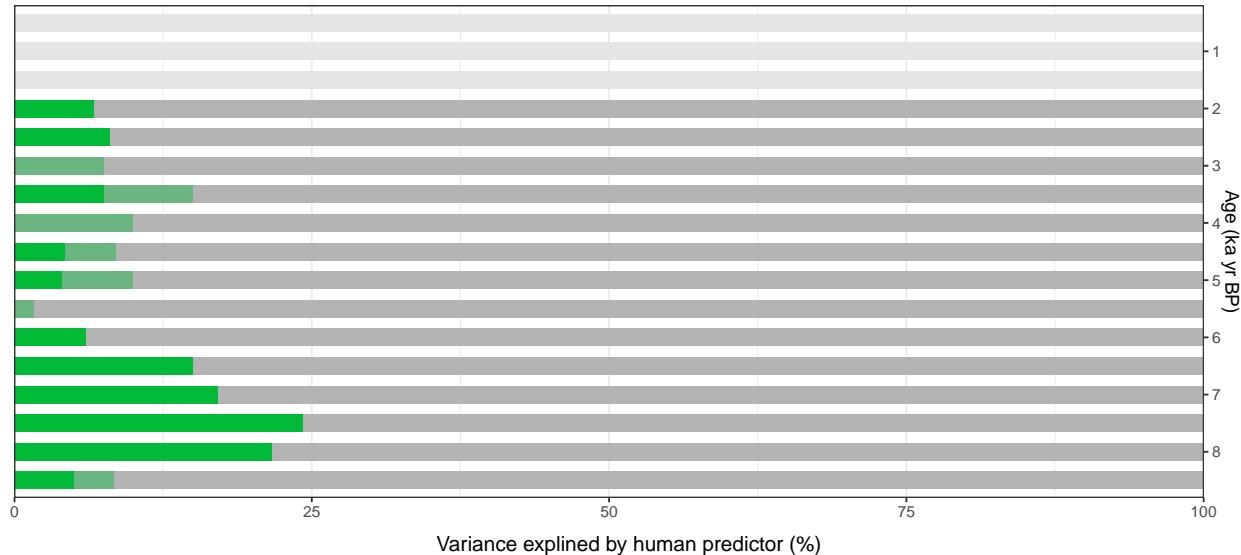
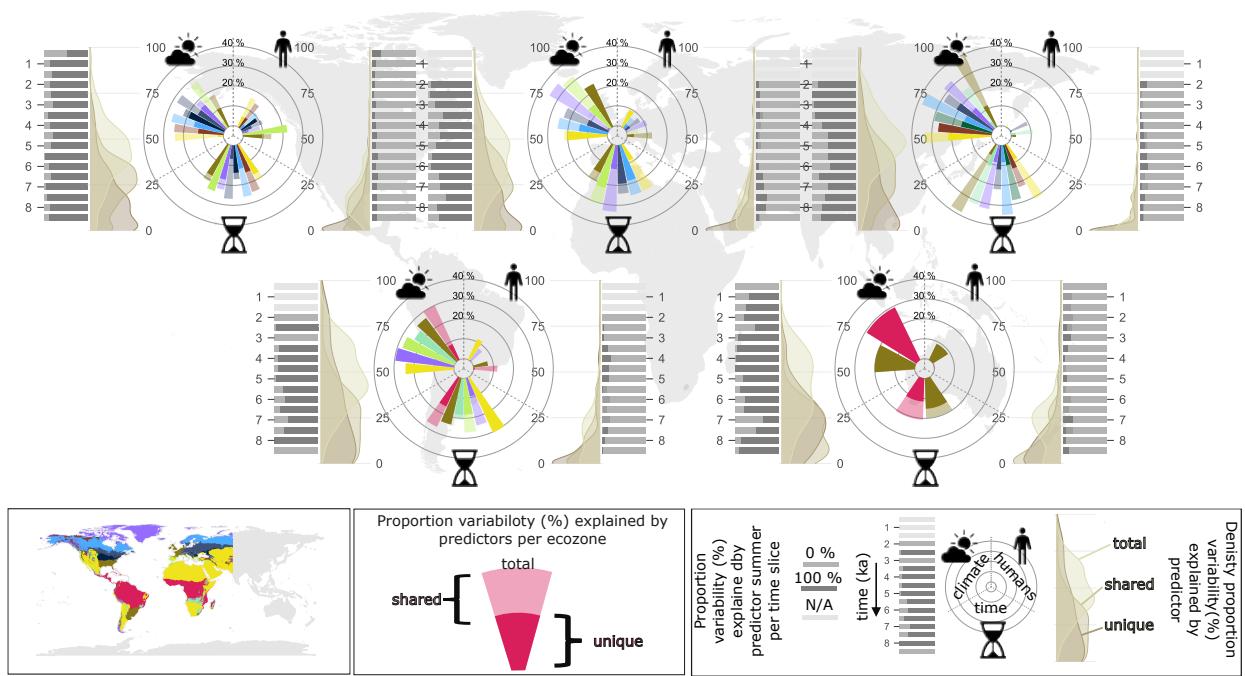


Figure 2

Finally, all those parts are combined in Figure 2.



Temporal variation constrained by past humans

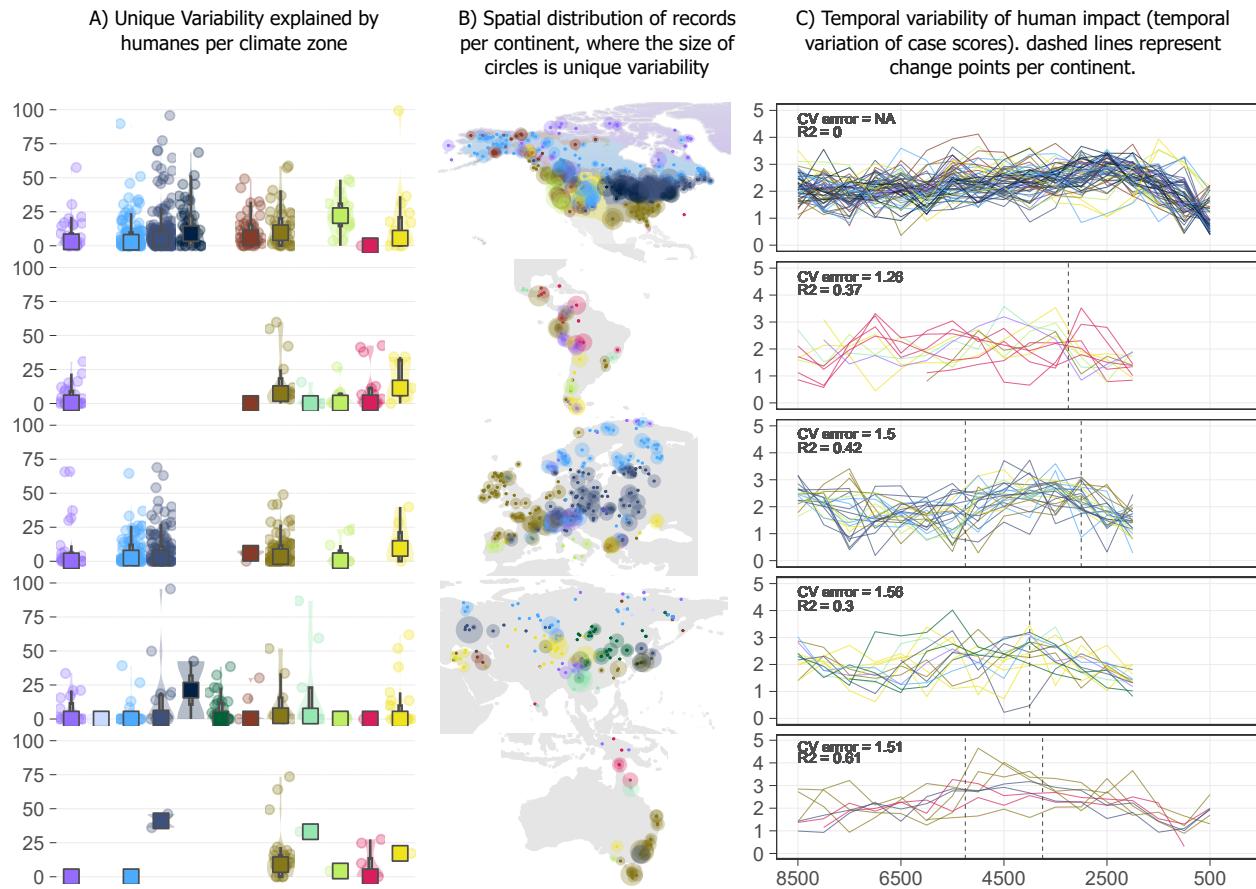
To visualise the variation of pollen assemblage properties constrained by humans, we made a partial constrained analysis with SPD as the predictor while partial out the variation by climate and time, and plotted the constrained axis with time. Only records with where past human explained more than 10% of the variation are shown. To test if there is a consistent changes in human impact (as measured by the PAPs) across the continents, we compiled a matrix of human impact for these sequences, and used this as a response matrix in a MRT analysis with age as the predictor. The fit of the data was not too bad, but the prediction error > 1 indicating low prediction power.

Here is an example for Europe:



Figure 3

Those are then combined with the distribution of variance explained unique by humans in Figure 3.

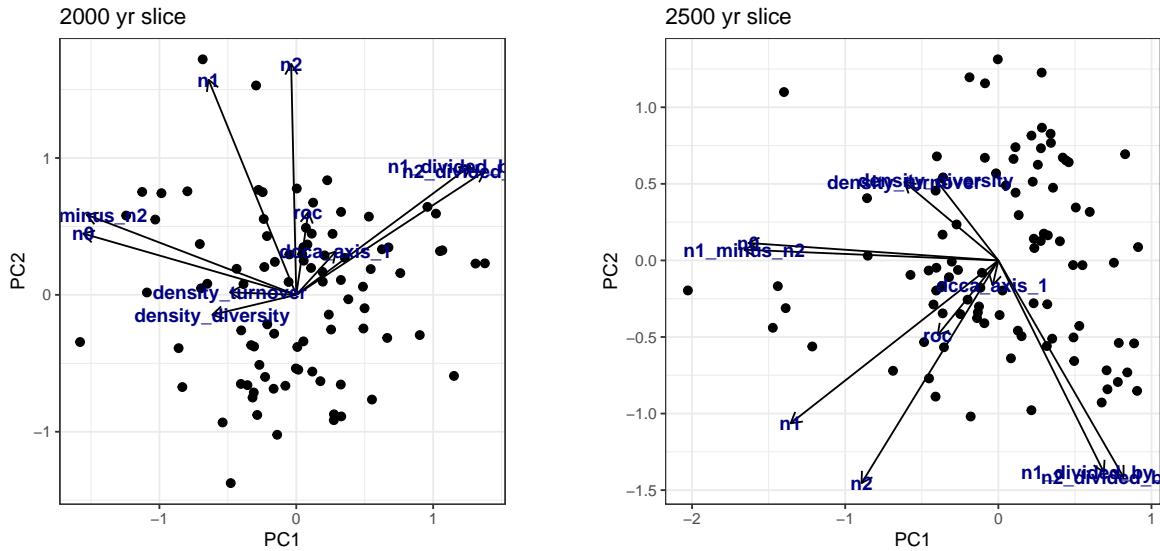


Numerical analysis of hypothesis 2

Multidimensional shift

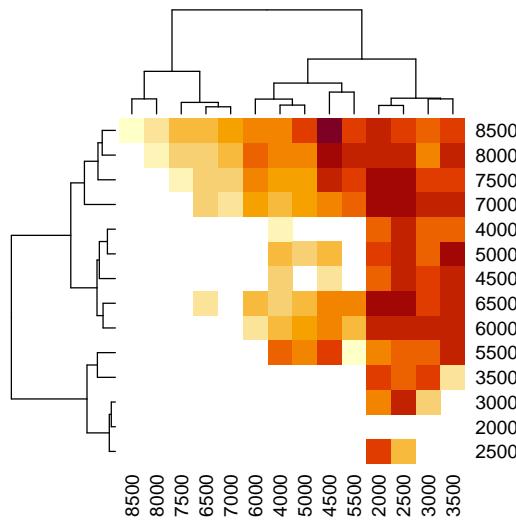
To test Hypothesis 2, that the interrelationships among pollen assemblage properties in response to humans, we integrated assemblage properties at each 500-year time slice within climate zones across continents. We used a principal component analysis (PCA) with standardisation of all indices.

Here is an example of PCA for Europe - Temperate Without dry season - 2000 yr slice:

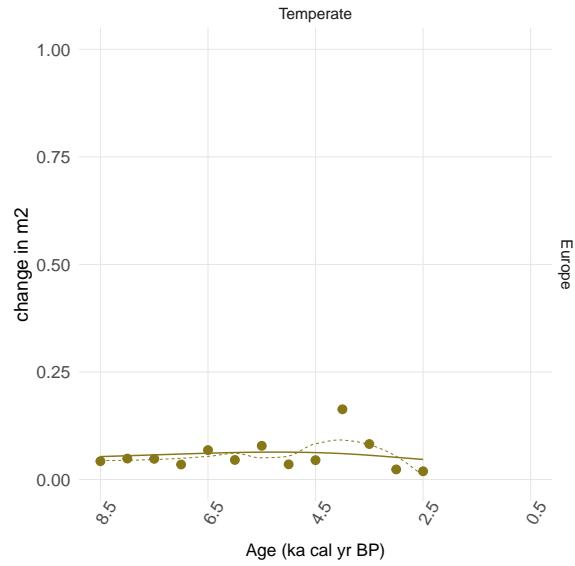


To quantify the overall changes in the interrelationship among pollen assemblage properties, we employed symmetrical Procrustes sum-of squares (m_2) to compare pairwise PCA changes within each time slice.

Here is a m_2 matrix of all time slices to each other for Europe - Temperate Without dry season:



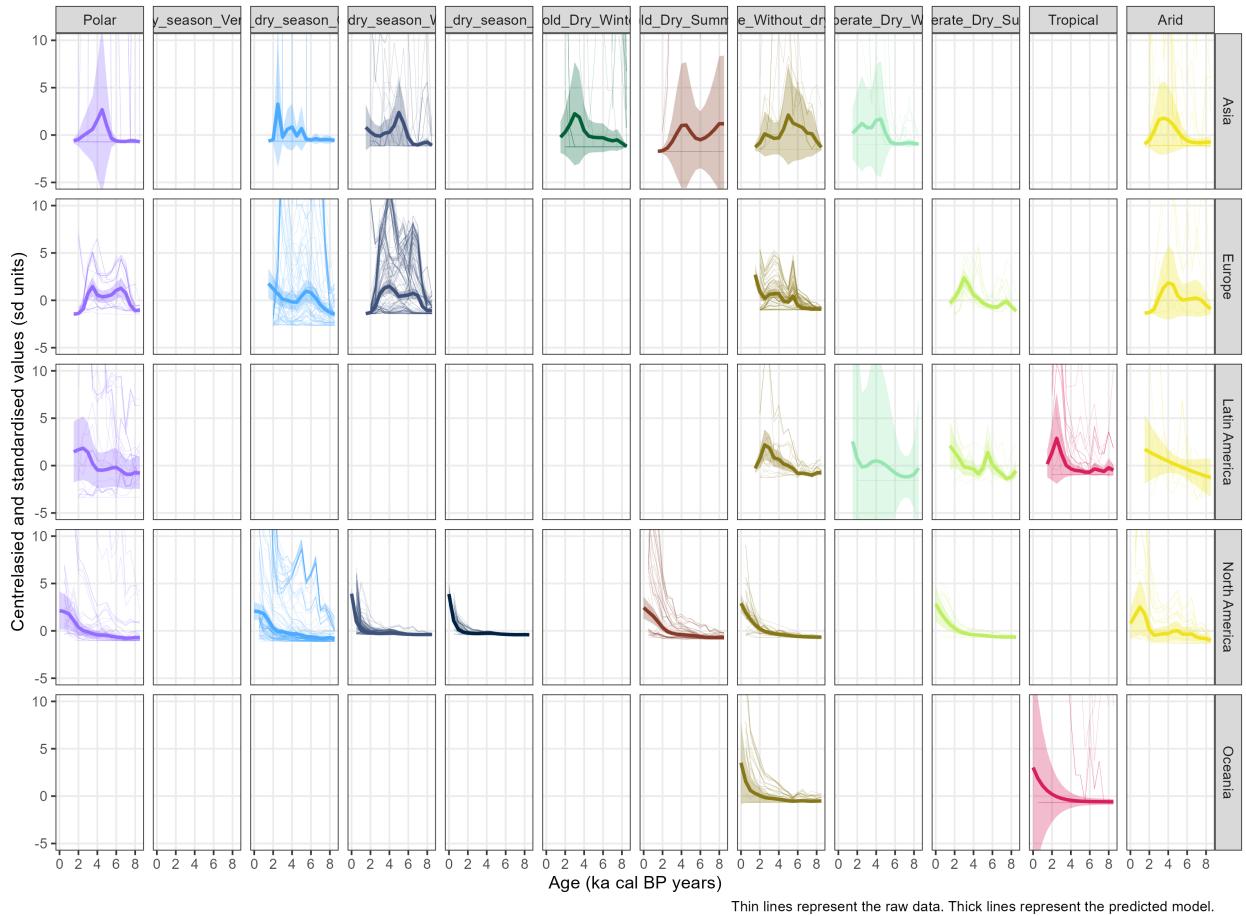
We can then plot the changes between subsequent time slices (diagonal+1 in the matrix) in relationship to the time. Here each point is a m_2 value between 2 time slices, the full line represent GAM model and the dashed line LOESS smoother. Here is an example for Europe - Temperate Without dry season:



General trends of predictors

Because m² matrix is on a level of climate zone within continent, we need to determine the general patterns of past humans (as represented by SPD) and palaeoclimate within the climate zones in the continents. We used hierarchical Generalised Additive Models (HGAM) to depict their main trend. We decided to use the climate zones as regional division, therefore the general trends within the climate zones are based on the data we have and not reconstructing new variables within the climate zones.

Here is a general trends of SPD (Extended Data Figure 8):



hVarPart

Subsequently, for each continent and climate zone, we used the pairwise m2 as distance matrices in hierarchical variation partitioning to quantify the importance of humans, climate, and time in explaining the multidimensional shifts among pollen assemblage properties through time. Subsequently, the explanatory variables derived from HGAMs are integrated into hierarchical variation partitioning to gauge the extent to which human activity and/or climate shifts influenced variations in the interrelationships among pollen assemblage properties.

The visualisation is the same as H1. Here is an example for Europe:

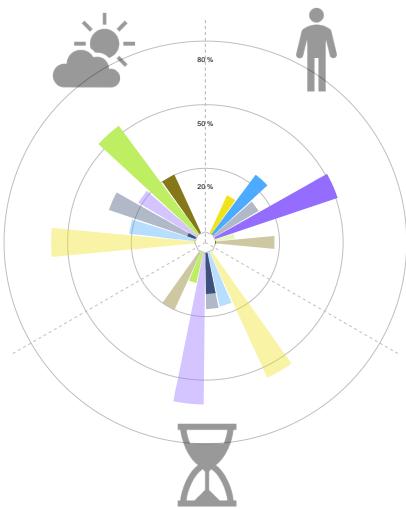


Figure 4

Finally, all those parts are combined in Figure 4.

Temporal trend of multidimensional shift of ecosystem properties (m²) per climate zone

Proportion variability (%) explained by predictors

