# **Project**

**Title:** The SPECIAL EPD Database: an expanded resource to document changes in vegetation and climate from pollen records from Europe, the Middle East and western Eurasia.

**Dates:** 2023-2024

**Funding organisations:** Funding for the development of the SPECIAL-EPD database was sourced from the Leverhulme Centre for Wildfires, Environment and Society through the Leverhulme Trust (grant: RC-2018-023).

# **Dataset**

**Title:** The SPECIAL EPD Database: an expanded resource to document changes in vegetation and climate from pollen records from Europe, the Middle East and western Eurasia.

**Summary description:** Pollen records from sedimentary archives can be used to reconstruct past changes in vegetation (Cruz-Silva et al., 2022; Cruz-Silva et al., 2024) and to make quantitative reconstructions of climate variables (e.g. Wei et al., 2021; Cruz-Silva et al., 2023; Liu et al., 2024). The SPECIAL-EPD is an expanded database of pollen records from Europe, the Middle East and western Eurasia. It builds on a pollen compilation covering the Middle East (EMBSecBIO database: Cordova et al., 2009; Harrison and Marinova, 2017; Marinova et al., 2017), data available from public data repositories (NEOTOMA: https://www.neotomadb.org/; PANGAEA: https://www.pangaea.de/) and data provided by the original authors for the Iberian peninsula (Liu et al., 2023) and other regions. The data have been extensively quality-controlled and mistakes have been corrected and documented. New BACON Bayesian age models, based on the recalibration of radiocarbon ages using INTCAL2020 (Reimer et al., 2020), are provided for all the records. These were constructed using the ‘rbacon’ R package (Blaauw et al., 2021) in the ‘AgeR’ R package (Villegas-Diaz et al., 2021). This package provides an optimum model for each record, based on the lowest quantified area between prior and posterior accumulation rate distribution curves (see Harrison et al., 2022). The pollen records have been taxonomically harmonised and the counts at this level are provided in the "clean" table. They have then been standardised at two further levels: one provides taxonomically harmonised counts at species level for woody plants and genus level for other plants ("intermediate" table); the second provides a further amalgamation of rare taxa into higher taxonomic groups (genus, sub-family, family as appropriate), after ensuring that this was consistent with their distribution in climate space ("amalgamated" table).

The SPECIAL-EPD database now contains 1758 entities from 1576 sites. New age models have been run for 1598 for the entities in total.

**Publication year:** 2024

**Creators:** Sandy P. Harrison, Mary-Ann Egbudom, Mengmeng Liu, Esmeralda Cruz-Silva (PhD student in Geography and Environmental Science, the University of Reading), Luke Sweeney (PhD student in Geography and Environmental Science, the University of Reading).

**Organisation:** Geography and Environmental Science, University of Reading, UK

**Rights Holders:** University of Reading, Esmeralda Cruz-Silva, and Luke Sweeney.

**Cite as:** [Harrison, Sandy](https://researchdata.reading.ac.uk/view/creators/Harrison=3ASandy=3A=3A.html), [Egbudom, Mary-Ann](https://researchdata.reading.ac.uk/view/creators/Egbudom=3AMary-Ann=3A=3A.html), [Liu, Mengmeng](https://researchdata.reading.ac.uk/view/creators/Liu=3AMengmeng=3A=3A.html), [Cruz-Silva, Esmeralda](https://researchdata.reading.ac.uk/view/creators/Cruz-Silva=3AEsmeralda=3A=3A.html) and [Sweeney, Luke](https://researchdata.reading.ac.uk/view/creators/Sweeney=3ALuke=3A=3A.html) (2024): The SPECIAL EPD Database: an expanded resource to document changes in vegetation and climate from pollen records from Europe, the Middle East and western Eurasia. University of Reading. Dataset. <https://doi.org/10.17864/1947.001295>

# **Terms of use**

This dataset is licensed under a Creative Commons Attribution 4.0 International Licence: <https://creativecommons.org/licenses/by/4.0/>. In order to assure traceability, any presentation, report, or publication that uses the SPECIAL-EPD should cite the dataset (https://doi.org/10.17864/1947.001295). If using individual sites, original literature citations provided in the database should also be cited.

# **Contents**

**Abstract:** The SPECIAL-EPD is an expanded database of pollen records from Europe, the Middle East and western Eurasia. It builds on a pollen compilation covering the Middle East (EMBSecBIO database: Cordova et al., 2009; Harrison and Marinova, 2017; Marinova et al., 2017), data available from public data repositories (NEOTOMA: https://www.neotomadb.org/; PANGAEA: https://www.pangaea.de/) and data provided by the original authors for the Iberian peninsula (Liu et al., 2023). It is designed to provide quality-controlled, taxonomically harmonised and standardised pollen counts with consistent age models for the reconstruction of past vegetation and climate.

**Structure and Contents of the Database**

The database tables are stored in two formats with similar structures.

A. Microsoft SQL Server (.sql file)

The database can be downloaded and accessed by importing into Microsoft SQL Server Management Studio. The following steps highlights how this can be implemented.

* Select Browse and locate SPECIAL\_EPD.sql file on the computer.
* Execute query
* Data will be imported and can be queried using T-SQL.

B. Comma-separated values (CSV) files

The database is also stored in separate csv files, each file represents a table. While these flat files are non-relational, they can be queried by associating shared columns using programs such as R packages and Python Pandas DataFrames. The relationships between the tables have been highlighted below, as well in Table 1 and Diagram 1.

The SPECIAL EPD database structure consists of several tables, each serving a specific purpose in storing and organizing data related to site metadata, samples, age models, dates, and taxon counts. Here is a summary description of each table:

1. Metadata:

- Columns: ID\_SITE, ID\_ENTITY, site\_name, entity\_name, latitude, longitude, elevation, site\_type, source, publication, doi, notes.

- Purpose: Contains metadata information about different sites and entities. A site can have multiple entities. ID\_ENTITY is unique, serving as a primary key.

2. Samples:

- Columns: ID\_ENTITY, ID\_SAMPLE, depth, thickness, chronology\_name, age\_type, age, age\_younger, age\_older, count\_type, sample\_type.

- Purpose: Stores information about individual samples, including their depth, thickness, age and chronology, and other age data as given originally. ID\_SAMPLE is unique, serving as a primary key.

3. Age Model:

- Columns: ID\_MODEL, ID\_SAMPLE, mean, median, UNCERT\_5, UNCERT\_25, UNCERT\_75, UNCERT\_95, model\_name.

- Purpose: Represents revised age models associated with specific entities, using a standard calibration and a standard age-modelling tool. ID\_SAMPLE is unique, serving as a primary key.

4. Dates:

- Columns: ID\_ENTITY, ID\_DATE\_INFO, date\_type, depth, thickness, lab\_num, age\_c14, age\_calib, error, material\_dated, age\_used, reason\_age\_not\_used, notes.

- Purpose: Contains information about different dates associated with a specific entity, including the type of date, the measurement and uncertainty, and what was dated. The table indicates whether dates were used in the original age model or are considered unreliable by the original authors.

5. Clean:

- Columns: ID\_SAMPLE, taxon, count.

- Purpose: Represents clean taxon count data associated with specific samples.

6. Intermediate:

- Columns: ID\_SAMPLE, taxon, count.

- Purpose: Stores intermediate taxon count data associated with specific samples.

7. Amalgamated:

- Columns: ID\_SAMPLE, taxon, count.

- Purpose: Contains amalgamated taxon count data associated with specific samples.

Queries and Implementation:

- The metadata table can be queried to obtain information about different sites and entities.

- Sample-related queries can be performed using the samples table, including details on depth, age, and sample characteristics.

- The age model table provides statistical information about age models associated with specific samples.

- The dates table can be used to retrieve information about dating methods, lab numbers, and associated errors.

- Clean, intermediate, and amalgamated tables have similar structures. They store taxon count data, under various taxon names, allowing researchers to analyse and compare pollen counts across different samples, without losing information on updated taxonomy.

This database structure facilitates the organization and retrieval of paleoecological data, supporting researchers in studying past environmental changes based on pollen information.



*Diagram 1: ER Diagram of EPD Tables.*

*Table 1. Summary of all the tables and their respective fields in the EPD.*

|  |  |  |
| --- | --- | --- |
| **Table** | **Field name** | **Definition** |
| metadata, dates, samples | ID\_ENTITY | Unique identifier for each entity |
| metadata | ID\_SITE | Unique identifier for each site |
| metadata | site\_name | Site name as given by original authors or as defined by us where there was no unique name given to the site |
| metadata | entity\_name | Name of entity, where an entity may be a separate core from the site or a separate type of measurement on the same core |
| metadata | latitude | Latitude of the entity, given in decimal degrees, where N is positive and S is negative |
| metadata | longitude | Longitude of the entity, given in decimal degrees, where E is positive and W is negative |
| metadata | elevation | Elevation of the sampling site, in metres above (+) or below (-) sea level |
| metadata | site\_type | Information about type of site (e.g. lake, peatland, terrestrial) |
| metadata | source | Source of pollen data |
| metadata | publication | The publication citation in bibtex format |
| metadata | doi | The digital object identifier (doi) for the pollen data |
| metadata | notes | Additional comments regarding an entity record |
| samples, age\_model, clean, intermediate, amalgamated | ID\_SAMPLE | Unique identifier for each pollen sample. |
| samples, dates | depth | Average sampling depth, in centimetres (sample table). Average depth where date was measured in centimetres (date\_info table). |
| samples | thickness | Sample thickness, in centimetres |
| samples | chronology\_name | The sample chronology used in the original publication |
| samples | age\_type | The unit type of the measured sample age (i.e. radiocarbon or calendar years). |
| samples | age | The age of the sample given in the original publication. |
| samples | age\_younger | The 5th percentile of the radiocarbon age |
| samples | age\_older | The 95th percentile of the radiocarbon age. |
| samples | count\_type | A type of count used in obtaining the number of pollen per sample. |
| samples | sample\_type | The type of sample (e.g. bog sediment) |
| dates | ID\_DATE\_INFO | Unique identifier for the date record |
| dates | date\_type | Technique used to obtain the date measurement |
| dates | thickness | Thickness of the sample used for dating, in centimetres |
| dates | lab\_num | Unique identifying code assigned by the dating laboratory |
| dates | age\_C14 | Uncalibrated radiocarbon age |
| dates | age\_calib | The calendar age of a date |
| dates | error | Analytical or measurement error on the date |
| dates | material\_dated | Material from which the date was obtained, if applicable |
| dates | age\_used | Indicates whether date was used by the author(s) in the construction of the original age model |
| dates | reason\_age\_not\_used | Indication of why a date was not used in the original age model. Blank if dates were used in original model |
| dates | notes | Additional comments regarding a date record |
| age\_model | mean | Mean age of the sample |
| age\_model | median | Median age of the sample |
| age\_model | UNCERT\_5 | Lower bound of the 5% confidence interval for the median age |
| age\_model | UNCERT\_25 | Lower bound of the 25% confidence interval for the median age |
| age\_model | UNCERT\_75 | Upper bound of the 75% confidence interval for the median age |
| age\_model | UNCERT\_95 | Upper bound of the 95% confidence interval for the median age |
| age\_model | model name | Age modelling technique used in the revised age calibration. |
| clean, intermediate, amalgamated | taxon | The name assigned to a particular type of pollen as given in the source (clean) or in various levels of standardisation (intermediate, amalgamated). |
| clean, intermediate, amalgamated | count | The number of pollen grains counted for each taxon. |

**References**

Blaauw, M., Christen, J. A., Lopez, M. A. A., Vazquez, J. E., Gonzalez V., O. M., Belding, T., Theiler, J., Gough, B., & Karney, C. (2021). *rba- con: Age-depth modelling using Bayesian statistics* (2.5.6) [R]. https:// CRAN.R-project.org/package=rbacon

Cordova, C.E., Harrison, S.P., Mudie, P.J., Riehl, S, Leroy, S.A.G., Ortiz, N., 2009. Pollen, plant macrofossil and charcoal records for palaeovegetation reconstruction in the Mediterranean-Black Sea Corridor since the Last Glacial Maximum. Quaternary International 197: 12-26.

Cruz-Silva, E., Harrison, S. P., Prentice, I. C., Marinova, E., Bartlein, P. J., Renssen, H., & Zhang, Y. (2023). Pollen-based reconstructions of Holocene climate trends in the eastern Mediterranean region. *Climate of the Past*, 19(11), 2093–2108. https://doi.org/10.5194/cp-19-2093-2023

Cruz-Silva, E., Harrison, S.P., Prentice, I.C., Marinova, E., 2024. Holocene vegetation dynamics of the Eastern Mediterranean region: Old controversies addressed by a new analysis. *Journal of Biogeography* 51: 294-310, DOI: 10.1111/jbi.14749

Cruz-Silva, E., Harrison, S.P., Marinova, E., Prentice, I.C., 2022. A new method based on surface-sample pollen data for reconstructing palaeovegetation patterns. *Journal of Biogeography* 49: 1381-1396, doi: 10.1111/jbi.14448.

Google Earth 9.0, (2017) *Europe 34°00’00.0”N to 73°00’00.0”N, 12°00'00.0"W to 45°00'00.0”E.* [Online] Available at: http://www.google.com/earth/index.html [Accessed 31 March 2023].

Harrison, S.P. and Marinova, E., 2017. EMBSeCBIO modern pollen biomisation. University of Reading Dataset.  [http://dx.doi.org/10.17864/1947.109](https://www.owamail.reading.ac.uk/owa/redir.aspx?C=948cH0RMKcW0GIgELrdxYAMR9QKsTLsGA8rx2tiwxMBcwKz8w1zUCA..&URL=http%3a%2f%2fdx.doi.org%2f10.17864%2f1947.109).

Harrison, S.P., Marinova, E., & Cruz-Silva, E. (2021). *EMBSeCBIO pollen database* [Data set]. University of Reading. https://doi.org/10. 17864/1947.309

Liu, M., Prentice, I.C., ter Braak, C.J.F, Harrison, S.P., 2020. An improved statistical approach for reconstructing past climates from biotic assemblages. *Proceedings of the Royal Society A, Mathematics* *A* 476: 20200346, 20200346, https://doi.org/10.1098/rspa.2020.0346

Liu, M., Shen, Y., González-Sampériz, P., Gil-Romera, G., ter Braak, C.J.F. Prentice, I.C., Harrison, S.P., 2023. Holocene climates of the Iberian Peninsula. *Climate of the Past* 19: 803-834, https://doi.org/10.5194/cp-19-803-2023.

Marinova, E., Harrison, S.P., Bragg, F., Connor, S., de Laet, V., Leroy, S., Mudie, P., Atanassova, J., Bozilova, E., Caner, H., Cordova, C., Djamali, M., Filipova-Marinova, M., Gerasimenko, N., Kouli, K., Kotthoff, U., Kvavadze, E., Lazarova, M., Novenko, E., Ramezani, E., Röpke, A., Shumilovskikh, L., Tantau, I., Tonkov, S., 2017. Pollen-derived biomes in the eastern Mediterranean-Black Sea-Caspian corridor. *Journal of Biogeography* 45: 484–499 DOI: 10.1111/jbi.13128

Reimer, P., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., 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 calibra- tion curve (0-55 cal kBP). *Radiocarbon*, *62*(4), 725–757. https://doi. org/10.1017/RDC.2020.41

Villegas-Diaz, R., Cruz-Silva, E., & Harrison, S. P. (2021). ageR: Supervised age models [R]. *Zenodo*, https://doi.org/10.5281/zenodo.4636716

Wei, D., González-Sampériz, P., Gil-Romera, G., Harrison, S.P., Prentice, I.C., 2021. Seasonal temperature and moisture changes in interior semi‐arid Spain from the last interglacial to the Late Holocene. *Quaternary Research,* 101, 143-155, doi:10.1017/qua.2020.108