The Indo-Pacific Pollen Database - a Neotoma constituent database

Annika V. Herbert[1,2], Simon G. Haberle[1,2], Suzette G.A. Flantua[3], Ondrej Mottl[3,4,5], Jessica L. Blois[6], John W. Williams[7], Adrian George[7], Geoff Hope[1†]

[1]Department of Archaeology and Natural History, Australian National University, Canberra, ACT 2601, Australia.

[2]Australian Research Council Centre of Excellence in Australian Biodiversity and Heritage, Australian National University, Canberra, ACT 2601, Australia.

[3]Department of Biological Sciences, University of Bergen, Bjerknes Centre for Climate Research, N-5020 Bergen, Norway.

[4]Center for Theoretical Study, Charles University, Jilská 1, 11000-CZ Praha 1, Czech Republic

[5]Department of Botany, Faculty of Science, Charles University, Benátská 2, CZ-12801 Prague, Czech Republic

[6]Department of Life and Environmental Sciences, University of California-Merced, Merced, CA 95343, USA.

[7]Department of Geography, University of Wisconsin, Madison, WI 53706, USA.

†Deceased

Correspondence to: Annika V. Herbert (annika.herbert@anu.edu.au)

Abstract. The Indo-Pacific Pollen Database (IPPD) is the brainchild of the late Professor Geoffrey Hope, who gathered pollen records from across the region to ensure their preservation for future generations of palaeoecologists. This noble aim is now being fulfilled by integrating the IPPD into the online Neotoma Palaeoecology Database, making this compilation available for public use. Here we explore the database in depth and suggest directions for future research. The IPPD comprises 226 fossil pollen records, most postdating 20 ka, but some extending as far back as 50 ka or further. Over 80% of the records are Australian, with a fairly even distribution between the different Australian geographical regions, the notable exception being Western Australia, which is only represented by 3 records. The records are also well distributed in modern climate space, the largest gap being in drier regions due to preservation issues. However, many of the records contain few samples or have fewer than 5 chronology control points, such as radiocarbon, luminescence or Pb-210 for the younger sequences. Average sedimentation rate for the whole database, counted as years per cm, is 64.8 yr/cm, with 61% of the records having a rate of less than 50 yr/cm. The highest sedimentation rate by geographical region occurs on Australia's east coast, while the lowest rates are from the Western Pacific. Overall, Australia has a higher

sedimentation rate than the rest of the Indo-Pacific region. The IPPD offers many exciting

35   research opportunities, such as examination of human impact on regional vegetation, contrasting first human arrival and colonisation, and assessment of rates of vegetation change during the Holocene. Merging the IPPD into Neotoma also facilitates inclusion of data from the Indo-Pacific region into global syntheses.

40   1. Introduction

The digital revolution and the advent of the internet transformed the landscape of fossil sample databases, enabling them to be shared with a global audience and heralding an era of scientific transparency and cooperation. The Neotoma Palaeoecology Database represents a significant stride in this direction, offering a diverse range of records including fossil pollen,

45   charcoal, vertebrates, diatoms, ostracods, insects and geochronological data (Williams *et al.,* 2018). This great resource has been used in hundreds of scientific publications to date, thus enabling the science of palaeoecology to take significant steps forward. In the era of open science, the importance of making scientific data publicly available cannot be overstated. Open access to data not only fosters transparency in research but also enables collaboration

50   and innovation across disciplines and geographical boundaries (Wolkovich et al., 2012; Record et al., 2022). This approach is crucial in the field of palaeoecology, where comprehensive and accessible databases are key to understanding ecological histories and predicting future environmental changes.

The journey of large databases in capturing fossil data spans several decades, beginning with

55   pioneering efforts like Newell's palaeontological database in 1952. The advent of spatial analyses of fossil data gave rise to significant developments, such as the European, North American and Latin American Pollen Databases, which were integral to the concept of a Global Pollen Database, all of which are now constituent databases in Neotoma (Gajewski, 2008; Whitmore et al., 2005; Fyfe et al., 2009; Flantua et al., 2015). A notable extension of

60   this endeavour was the Indo-Pacific Pollen Database (IPPD), originally compiled as part of the BIOME 6000 project. The aim of this project was to create biome maps for various periods (Prentice and Webb, 1998; Pickett et al., 2004), namely the present day, the mid-Holocene, defined as $6,000 \pm 500$ cal yr BP and the Last Glacial Maximum (LGM), defined as $21,000 \pm 1,000$ cal yr BP (Pickett *et al.,* 2004). Subsequent enhancements involved the

65   inclusion of full records, as opposed to time slices, with a focus on Australian records

(Herbert and Harrison, 2016), the full incorporation of an Indo-Pacific compilation done by the late Professor Geoff Hope (Hope et al., 1999), and modern samples (<100 yr BP). The latter served as a training data set for palaeoclimate reconstructions for comparison with palaeoclimate models (Herbert and Harrison, 2016). This version of the database has been

70   used in multiple wide-ranging studies to date, including global overviews of rates of change processes (Mottl et al., 2021) and pollen taxonomic harmonisation processes (Birks *et al.,* 2023). Here we explore this new constituent database in depth and suggest directions for future research.

75   ## 2. Methods

Due to the recognised importance of openly accessible palaeoecological databases, fully integrating the IPPD into Neotoma (https://www.neotomadb.org/) has been a long-term goal. Since 2021 a concerted effort has been made to fulfil this goal, which has involved getting hundreds of records in the right format and double-checking all the data and metadata against

80   available publications, as well as removing duplicated entries and other data entry errors. Being such an underrepresented region, it also involved adding over 400 pollen taxa to Neotoma's taxonomic structure. This in itself is a complicated procedure, as every single taxon needs to be validated against up-to-date and trusted sources. For most of these taxa, we used The Australian Plant Name Index

85   (https://biodiversity.org.au/nsl/services/search/names), and references therein.

Age models for all records have been reconstructed using Bacon (Blaauw and Christen, 2011) in R (R Core Team, 2022), versions depending on when the age model was constructed, as the database has been updated in stages. Radiocarbon dates were calibrated using either SHCal20 (Hogg et al., 2020) or SHCal13 (Hogg et al., 2013), again depending on when the

90   age model was constructed.

For ease of presentation, we will here only present fossil records with more than 2 chronological control points and at least 3 stratigraphic levels, thus excluding many surface sample collections. The minimum number of chronological control points was chosen to be able to construct robust age depth models. Chronological control points include dates

95   obtained through radiocarbon dating, luminescence dating (either thermo or optically

stimulated), U/Th, or Pb-210 for younger sequences. Plots were produced in R (R Core Team, 2022) to show different aspects of the IPPD (see supplementary information).

Modern observational climate data was taken from the CRU TS v4.07 gridded dataset at 0.5° grid cell resolution (Harris et al., 2020). The location of each record in the IPPD was then

100 plotted against the full modern dataset, with the chosen climate variables being: Mean Annual Precipitation (MAP), Mean Annual Temperature (MAT), Mean Temperature of the Coldest month (MTCO) and Mean Temperature of the Warmest Month (MTWA).

3. Results and discussion

105 The IPPD holds records from a total of 530 sites, many of which contain only surface samples. A total of 226 records covering the Indo-Pacific region are presented here, which excludes surface samples and poorly dated records. These 226 records contain 9765 samples, with an average of 43 samples per record (SD: 51.3). There are 2461 unique taxa in the database, with an average of 56 taxa per record (SD: 31.9).

110 *3.1 Spatial distribution*

Over 83% of the records are from Australia, with high representation from each Australian region, apart from Western Australia (Fig. 1a). The majority of pollen samples available through the IPPD consist of raw counts (58.4%, Fig. 1b). Just over a quarter of the records (27%) have had to be digitised from publications or theses, due to the raw data no longer

115 being available. Every region in the IPPD is represented mostly by sequences containing raw counts, which can therefore be used to verify the quality of the sequences from digitised sources from the same region (Figs. 1b, A2). There is also a good representation in modern climate space, with the IPPD sites covering most of the available observed climate space, the largest gap being in the driest regions due to preservation issues (Fig. 2). While there is a

120 high number of sites from the southeastern coast of Australia (Appendix A, Fig. A1), the sites in the IPPD cover much of the available modern climate space despite this spatial bias (Fig. 2). This is invaluable when performing palaeoclimate reconstructions, as most methods use some form of modern analogue calibration (Herbert and Harrison, 2016).

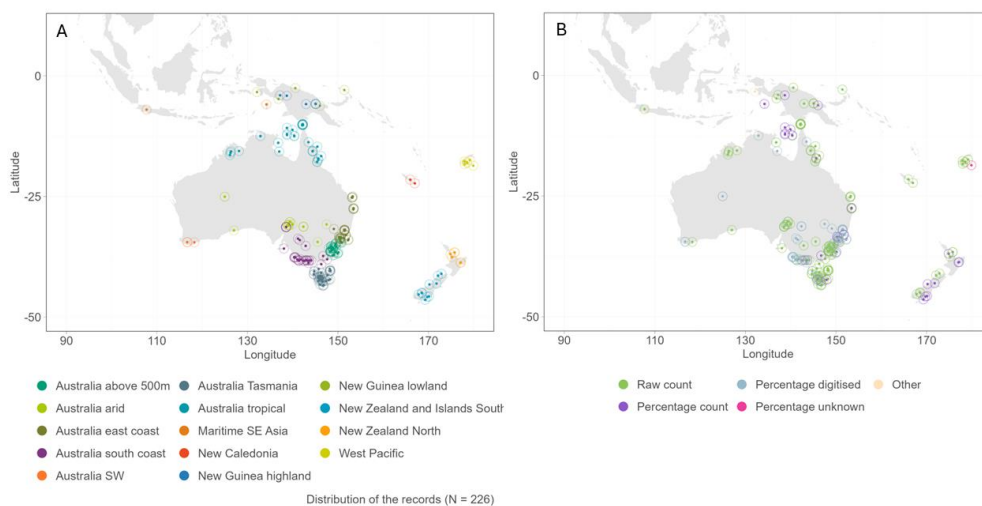



Figure 1, a) Site locations by geographical region, b) geographical distribution of records by count type. For detailed histograms, see Fig. A1 and A2 in the Appendix.

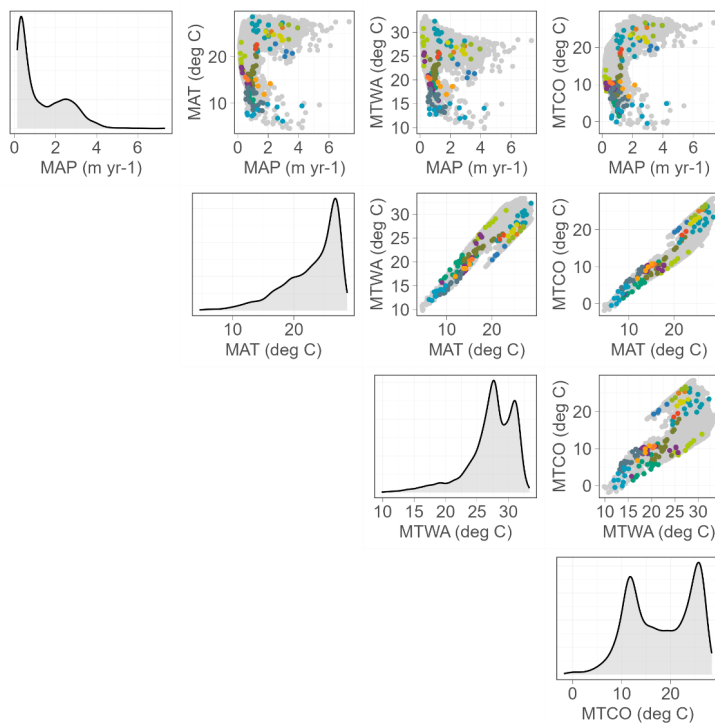



Figure 2, Modern climate space covered by sites in the IPPD. Available observed climate
space for the Indo-Pacific region in grey (from CRU TS v4.07, Harris *et al.,* 2020), IPPD
130   sites coloured according to geographical region, as presented in Fig. 1a. The line diagrams
show the distribution of each climate variable.

3.2 Depositional environment

There are a total of 33 different depositional environments represented in the database, with
lakes and wetlands being the main source of records (73% of sites) (Fig. 3). 31% of the
135   records come from lakes, natural or man-made, with a further 42% coming from wetlands,
such as swamps or fens.

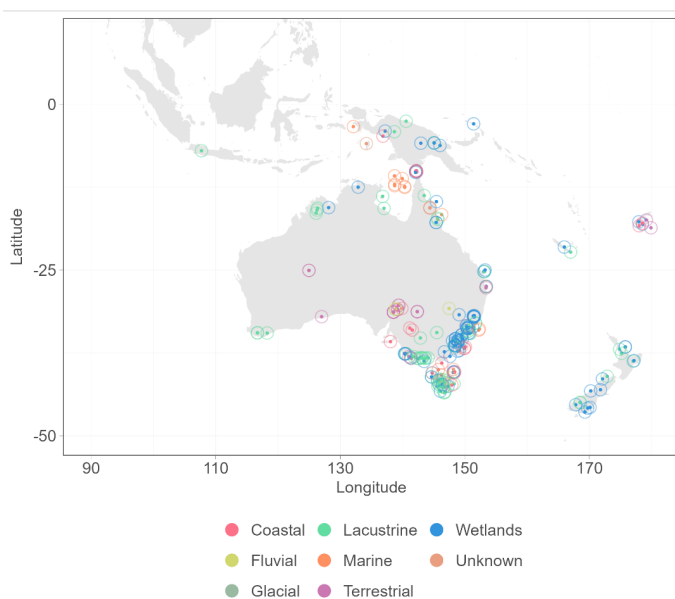
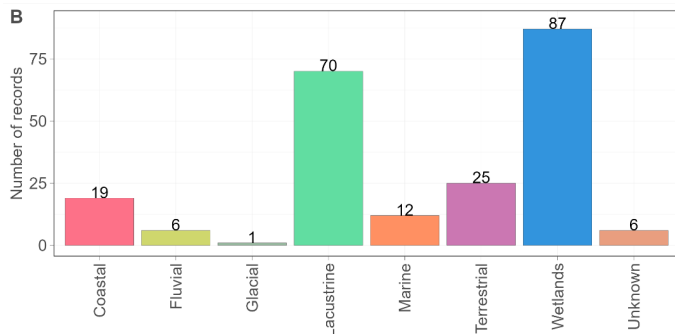



Figure 3, a) geographical distribution of records by depositional environment; b) number of records by depositional environment, binned into broader groupings. For detailed histograms, see Fig. A3 in the Appendix. For a full list of depositional environments, see Supplementary Table S1.

3.3 Temporal distribution and chronological control

There are a total of 1637 chronological control points in the database, with an average of 7 points per record (SD: 7.8). 52.6% of the records contain fewer than 5 chronological points each, meaning they may be poorly dated (Fig. 4a), depending on the number of levels associated with the site (Fig. 4b).

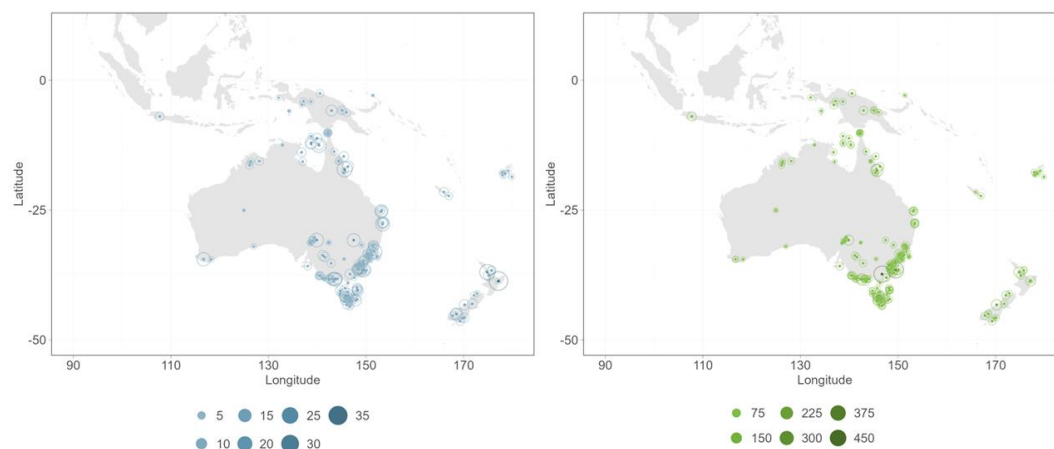



Figure 4, a) Geographical distribution of sites by number of chronological control points, b) geographical distribution of sites by number of levels. For a detailed histogram, see Fig. A4 and A5 in the Appendix.

Some records (17 out of 226, e.g. 7.5%) in the IPPD date back more than 50 ka BP, but most of them are younger than 20 ka (180 out of 226, e.g. 79.6%). In addition, many sequences are not continuous and are composed of a few levels at the older end (Fig. 5). No clear geographical pattern in the distribution of short sequences or sequences with fewer than 5 chronological control points can be discerned from Fig. 4, meaning the length of the

sequences is not necessarily climate dependent. However, most of the older sequences are located close to the coast where they are more likely to receive high rainfall, keeping the sediments wet and preserving the pollen grains (Fig. 5). Keeping the sediments waterlogged also decreases the risk of erosion, thereby benefiting the development of long sequences of 160 well-preserved pollen (Delcourt and Delcourt, 1980; Lowe, 1982).

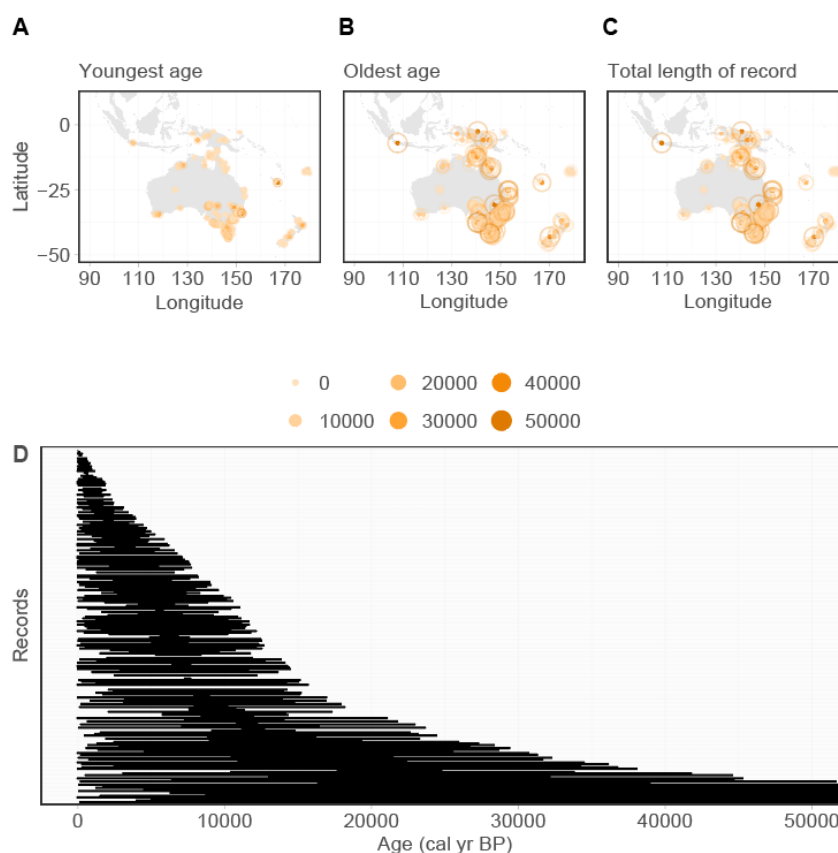



Figure 5, a-c) Geographical distribution of records by age, d) age distribution of the records.

3.4 Sedimentation rate

For 61% of the records, it takes less than 50 years to accumulate 1 cm of sediment. For 1.3% 165 of the records, it takes 300 years or more (Fig. 6 and Appendix A, Fig. A6). The average sedimentation rate for the IPPD is 64.8 yr/cm, but there is a very wide range, with the slowest regional sedimentation rate (largest number of years/cm), on average being from the east coast of Australia. The region with the largest range of values is arid Australia, encompassing

the highest and lowest rates overall. All regions in Australia bar one (south coast) has a

170   slower average rate than any non-Australian region, and also a much larger range (Fig. 6b). This could be due to lower average rainfall in Australia, which may inhibit biological activity in the lakes and wetlands that represent the majority of sites in the IPPD and slow down sediment accumulation. This is known to be an issue in arid and semi-arid areas (Ward and Larcombe, 2003), and with the rest of the regions represented in the IPPD being generally

175   high rainfall regions, this seems a likely explanation for the differences in rates. The average annual rainfall across all Australian IPPD sites is 1010 mm yr^{-1}, and across all non-Australian IPPD sites 2505 mm yr^{-1}. The reason Australian regions have larger ranges is probably due to the number of sites represented, with most sites in the database being Australian. The least represented Australian region, the southwest, also has the smallest range of any Australian

180   region.

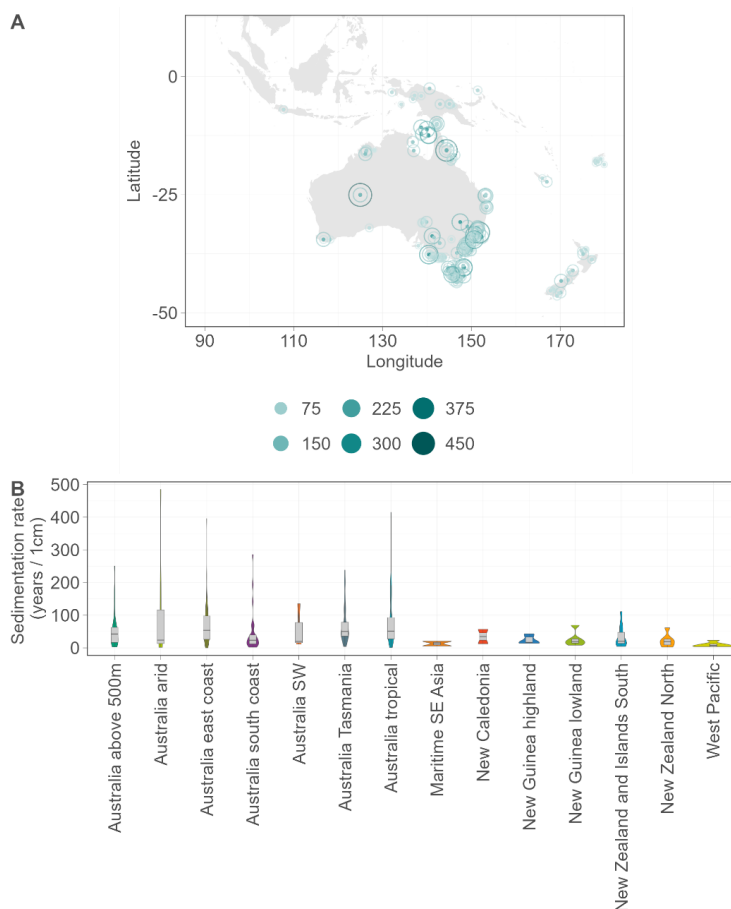



Figure 6, a) Geographical distribution of records by sedimentation rate; b) sedimentation rate by geographical region, with the mean marked by a line. For a detailed histogram of the number of records by sedimentation rate, see Fig. A6 in the Appendix.

185   A large study of sedimentation rates in eastern North America found that on average mid-latitude (40-50°N) sites accumulated sediments faster (low number of years/cm) than low- or high-latitude sites (Webb and Webb, 1988). The authors attributed the rate in low-latitude sites to the fact that most of these sites were shallow and affected by dry seasons, making them prone to periods of erosion. This could certainly also be the case for many of the

190   Australian sites, as it is common for lakes here to fluctuate in size or to periodically dry out completely. However, due to different methods of calculating sedimentation rates, no direct comparison can be made to other large-scale studies.

4. Conclusions and future work

The Indo-Pacific region has been poorly represented in global repositories in the past

195   (Gajewski, 2008). Having this resource available through Neotoma for the first time will go some way to create a truly global pollen database, and full integration of the IPPD into Neotoma complements other efforts focusing on getting more datasets into Neotoma from Africa (Lézine et al., 2021) and South America (Flantua et al., 2015). Neotoma operates with a constituent database structure (Williams et al. 2018) and sites within the IPPD can be

200   viewed through Neotoma's informatics ecosystem (e.g., through the Explorer app (https://apps.neotomadb.org/explorer/) or using the *neotoma* R package (Dominguez Vidaña and Goring 2023)). Going forward, Annika Herbert will serve as lead steward for the IPPD within Neotoma. Possible future analyses of our database include the examination of human impact on regional vegetation, contrasting first human arrival and colonisation (e.g. López-

205   Sáez et al., 2014; Flantua et al., 2016), examining human cultures and food production based on anthropogenic indicators (e.g. Flantua and Hooghiemstra, 2023) or the assessment of rates of vegetation change during the Holocene (e.g. Mottl et al., 2021). With such a large database now publicly accessible and open access workflows to process and standardise large compilations (Flantua et al., 2023; Vidaña and Goring, 2023), countless opportunities for

210   further study are available, including global and hemispheric syntheses.

Acknowledgements

The late Professors Geoff Hope and Eric Grimm were integral in the early stages of this project as well as in creating earlier versions of the database and initial upload efforts. None

215 of the current work would be possible without their efforts. Data were obtained from the Neotoma Paleoecology Database (http://www.neotomadb.org) and the IPPD constituent database. The work of data contributors, data stewards, and the IPPD / Neotoma community is gratefully acknowledged. The assistance of Matthew Jacques, Matthew Langer, Michael Rehani, Grace Roper, and Jocelyn Wai-Yee Lam in preparing IPPD records for upload as

220 undergraduate students at University of Wisconsin–Madison is gratefully acknowledged.

FUNDING

AVH is supported by the Australian Research Council's Centre of Excellence in Biodiversity and Heritage. OM and SGAF have been supported by the European Research Council (ERC)

225 under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 741413) to the project called 'HOPE Humans On Planet Earth—Long-term Impacts on biosphere dynamics'. Additionally, SGAF acknowledges support from Trond Mohn Research Foundation (TMF) and the University of Bergen for the startup grant 'TMS2022STG03' to the project called 'Past, Present, and Future of Alpine Biomes

230 Worldwide - PPF-Alpine'. OM has been supported by the Czech Science Foundation PIF grant (23-063861), by the Charles University Research Centre program (UNCE/24/SCI/006), and by the Institutional Support for Science and Research of the Ministry of Education, Youth and Sports of the Czech Republic. JLB was supported by the U.S. National Science Foundation Division of Earth Sciences (NSF EAR) 1948579.

235

DATA AVAILABILITY

The code for all figures is available at our public Github repo here; https://github.com/HOPE-UIB-BIO/IPPD_overview. All data were processed in R using Fossilpol (Flantua et al., 2023). Over half the sites are freely available through Neotoma, the rest are in the process of

240 being uploaded, to be completed by April 2025.

AUTHOR CONTRIBUTIONS

AH, SG, SF and OM conceptualised the paper; AH wrote the first draft and performed data analysis; OM made the figures; all authors except GH helped with data processing and

245 reviewed drafts. GH compiled the first database and initialised the project.

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

250 # References

Birks, H.J.B., Bhatta, K.P., Felde, V.A., Flantua, S.G.A., Mottl, O., Haberle, S.G., Herbert, A., Hooghiemstra, H., Birks, H.H., Grytnes, J.-A., Seddon, A.W.R.: Approaches to pollen taxonomic harmonisation in Quaternary palynology. Rev. Palaeobot. Palyno., 319, 104989, https://doi.org/10.1016/j.revpalbo.2023.104989, 2023.

255 Blaauw, M. and Christen, J.A.: Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Anal., 6 (3), 457-474. DOI: 10.1214/11-BA618, 2011.

Delcourt, P.A., Delcourt, H.R.: Pollen preservation and Quaternary environmental history in the southeastern United States. Palynology, 4, 215-231,
260 https://doi.org/10.1080/01916122.1980.9989209, 1980.

Flantua, S.G.A., Hooghiemstra, H.: Anthropogenic pollen indicators: Global food plants and Latin American human indicators in the pollen record. Scientific Data 10, 721. doi.org/10.1038/s41597-023-02613-1 2023.

Flantua, S.G.A., Hooghiemstra, H., Grimm, E.C., Behling, H., Bush, M.B., González-
265 Arango, C., Gosling, W.D., Ledru, M.-P., Lozano-García, Maldonado, A., Prieto, A.R., Rull, V., Van Boxel, J.H.: Updated site compilation of the Latin American Pollen Database. Rev. Palaeobot. Palyno. 223, 104-115. doi.org/10.1016/j.revpalbo.2015.09.008 2015.

Flantua, S.G.A., Hooghiemstra, H., Vuille, M., Behling, H., Carson, J.F., Gosling, W.D., Hoyos, I., Ledru, M.P., Montoya, E., Mayle, F., Maldonado, A., Rull, V., Tonello, M.S.,
270 Whitney, B.S., González-Arango, C.: Climate variability and human impact in South

America during the last 2000 years: synthesis and perspectives from pollen records. Clim. Past 12, 483-523. doi:10.5194/cp-12-483-2016 2016.

Flantua, S.G.A., Mottl, O., Felde, V.A., Bhatta, K.P., Birks, H.H., Grytnes, J.-A., Seddon, A.W.R., Birks, H.J.B.: A guide to the processing and standardization of global
275   palaeoecological data for large-scale syntheses using fossil pollen. Global Ecol. Biogeogr. 32, 1377-1394. DOI: 10.1111/geb.13693 2023.

Fyfe, R.M., de Beaulieu, J.-L., Binney, H., Bradshaw, R.H.W., Brewer, S., Le Flao, A., Finsinger, W., Gaillard, M.-J., Giesecke, T., Gil-Romera, G., Grimm, E.C., Huntley, B., Kunes, P., Kühl, N., Leydet, M., Lotter, A.F., Tarasov, P.E., Tonkov, S.: The European
280   Pollen Database: past efforts and current activities. Veg. Hist. Archaeobot., 18, 417-424, https://doi.org/10.1007/s00334-009-0215-9, 2009.

Gajewski, K.: The Global Pollen Database in biogeographical and palaeoclimatic studies. Prog. Phys. Geog., 32 (4), 379-402, DOI: 10.1177/0309133308096029, 2008.

Herbert, A.V., Harrison, S.P.: Evaluation of a modern-analogue methodology for
285   reconstructing Australian palaeoclimate from pollen Rev. Palaeobot. Palyno., 226, 65-77, http://dx.doi.org/10.1016/j.revpalbo.2015.12.006, 2016.

Hogg, A.G., Hua, Q., Blackwell, P.G., Niu, M., Buck, C.E., Guilderson, T.P., Heaton, T.J., Palmer, J.G., Reimer, P.J., Reimer, R.W., Turney, C.S.M., Zimmerman, S.R.H.: SHCal13 Southern Hemisphere calibration, 0-50,000 years cal BP. Radiocarbon, 55 (4), 1889-1903.
290   doi:10.2458/azu_js_rc.55.16783, 2013.

Hogg, A.G., Heaton, T.J., Hua, Q., Palmer, J.G., Turney, C.S.M., Southon, J., Bayliss, A., Blackwell, P.G., Boswijk, G., Bronk Ramsey, C., Pearson, C., Petchey, F., Reimer, P., Reimer, R., Wacker, L.: SHCal20 Southern Hemisphere calibration, 0-55,000 years cal BP. Radiocarbon, 62 (4), 759-778. doi: 10.1017/RDC.2020.59, 2020.

295   Hope, G.S., O'Dea, D., Southern, W.: Holocene vegetation histories in the Western Pacific – alternative records of human impact, in: Le Pacifique de 5000 à 2000 avant le present: suppléments à l'histoire d'une colonisation, edited by: Lilley, I., Galipaud, J.-C., IRD, Paris, France, 387-404, 1999.

Lézine, A.-M., Ivory, S.J., Gosling, W.D., Scott, L.: The African Pollen Database (APD) and
300   tracing environmental change: State of the Art, in: Quaternary Vegetation Dynamics, edited
by: Runge, J., Gosling, W., Lézine, A.-M., Scott, L., CRC Press, London, UK, 5-12,
doi.org/10.1201/9781003162766, 2021.

López-Sáez, J.A., Abel-Schaad, D., Pérez-Díaz, S., Blanco-González, A., Alba-Sánchez, F.,
Dorado, M., Ruiz-Zapata, B., Gil-García, M.J., Gómez-González, C., Franco-Múgica, F.:
305   Vegetation history, climate and human impact in the Spanish Central System over the last
9000 years. Quatern. Int. 353, 98-122., doi.org/10.1016/j.quaint.2013.06.034, 2014.

Lowe, J.J.: Three Flandrian pollen profiles from the Teith Valley, Perthshire, Scotland. II.
Analysis of deteriorated pollen. New Phytol., 90, 371-385, https://doi.org/10.1111/j.1469-
8137.1982.tb03268.x, 1982.

310   Mottl, O., Flantua, S.G.A., Bhatta, K.P., Felde, V.A., Giesecke, T., Goring, S., Grimm, E.C.,
Haberle, S., Hooghiemstra, H., Ivory, S., Kunes, P., Wolters, S., Seddon, A.W.R., Williams,
J.W.: Global acceleration in rates of vegetation change over the past 18,000 years. Science,
372 (6544), 860-864, DOI: 10.1126/science.abg1685, 2021.

Newell, N.D.: Periodicity in invertebrate evolution. J. Paleontol., 26, 371-385, 1952.

315   Pickett, E.J., Harrison, S.P., Hope, G., Harle, K., Dodson, J.R., Kershaw, A.P., Prentice, I.C.,
Backhouse, J., Colhoun, E.A., D'Costa, D., Flenley, J., Grindrod, J., Haberle, S., Hassell, C.,
Kenyon, C., Macphail, M., Martin, H., Martin, A.H., McKenzie, M., Newsome, J.C., Penny,
D., Powell, J., Raine, J.I., Southern, W., Stevenson, J., Sutra, J.-P., Thomas, I., van der Kaars,
S., Ward, J.: Pollen-based reconstructions of biome distributions for Australia, Southeast Asia
320   and the Pacific (SEAPAC region) at 0, 6000 and 18,000 ^{14}C yr BP. J. Biogeogr., 31, 1381-
1444, https://doi.org/10.1111/j.1365-2699.2004.01001.x, 2004.

Prentice, I.C., Webb, T.: BIOME 6000: reconstructing global mid-Holocene vegetation
patterns from palaeoecological records. J. Biogeogr., 25, 997-1005,
https://doi.org/10.1046/j.1365-2699.1998.00235.x, 1998.

325   Prentice, I.C., Guiot, J., Huntley, B., Jolly, D., Cheddadi, R.: Reconstructing biomes from
palaeoecological data: a general method and its application to European pollen data at 0 and 6
ka. Clim. Dynam., 12, 185-194, https://doi.org/10.1007/BF00211617, 1996.

R Core Team: R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. https://www.r-project.org/, 2022.

330   Record, S., Jarzyna, M.A., Hardiman, B., Richardson, A.D.: Open data facilitate resilience in science during the COVID-19 pandemic. Front. Ecol. Environ., 20 (2), 76, doi:10.1002/fee.2468, 2022.

Vidaña, S.D., Goring, S.J.: neotoma2: An R package to access data from the Neotoma Paleoecology Database. Journal of Open Source Software, 8 (91), 5561,

335   doi.org/10.21105/joss.05561, 2023.

Ward, I., Larcombe, P.: A process-orientated approach to archaeological site formation: application to semi-arid Northern Australia. J. Archaeol. Sci., 30, 1223-1236, https://doi.org/10.1016/S0305-4403(02)00253-4, 2003.

Webb, R.S., Webb, T.: Rates of sediment accumulation in pollen cores from small lakes and

340   mires of eastern North America. Quaternary Res., 30, 284-297, doi:10.1016/0033-5894(88)90004-X, 1988.

Whitmore, J., Gajewski, K., Sawada, M., Williams, J.W., Shuman, B., Bartlein, P.J., Minckley, T., Viau, A.E., Webb III, T., Shafer, S., Anderson, P., Brubaker, L.: Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications.

345   Quaternary Sci. Rev., 24, 1828-1848, doi:10.1016/j.quascirev.2005.03.005, 2005.

Williams, J.W., Grimm, E.C., Blois, J.L., Charles, D.F., Davis, E.B., Goring, S.J., Graham, R.W., Smith, A.J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A.C., Betancourt, J.L., Bills, B.W., Booth, R.K., Buckland, P.I., Curry, B.B., Giesecke, T., Jackson, S.T., Latorre, C., Nichols, J., Purdum, T., Roth, R.E., Stryker, M., Takahara, H.: The Neotoma

350   Paleoecology Database, a multiproxy, international, community-curated data resource. Quaternary Res., 89, 156-177, doi:10.1017/qua.2017.105, 2018.

Wolkovich, E.M., Regetz, J., O'Connor, M.I.: Advances in global change research require open science by individual researchers. Global Change Biol., 18 (7), 2102-2110, doi.org/10.1111/j.1365-2486.2012.02693.x, 2012.

355

Appendix A

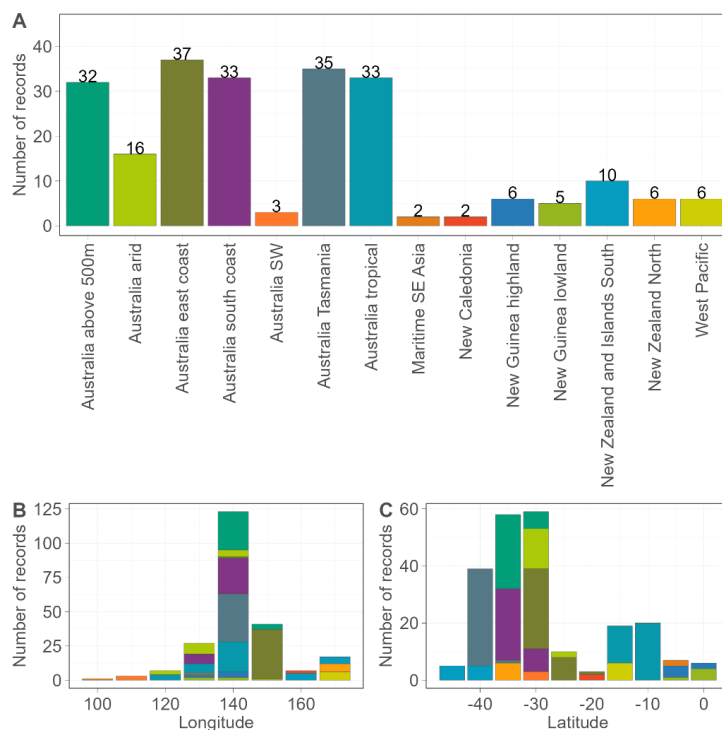



Figure A1, a) Number of records by geographical region; b) Number of records by latitude, colours representing the regions from a; c) Number of records by longitude, colours representing the regions from a.

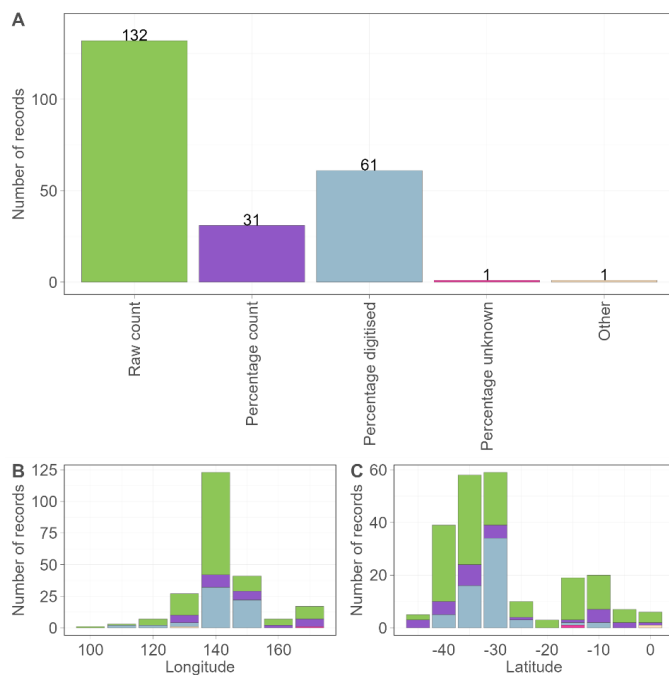



Figure A2, a) Number of records by count type; b) Number of records by latitude, colours representing count type from a; c) Number of records by longitude, colours representing count type from a.

365

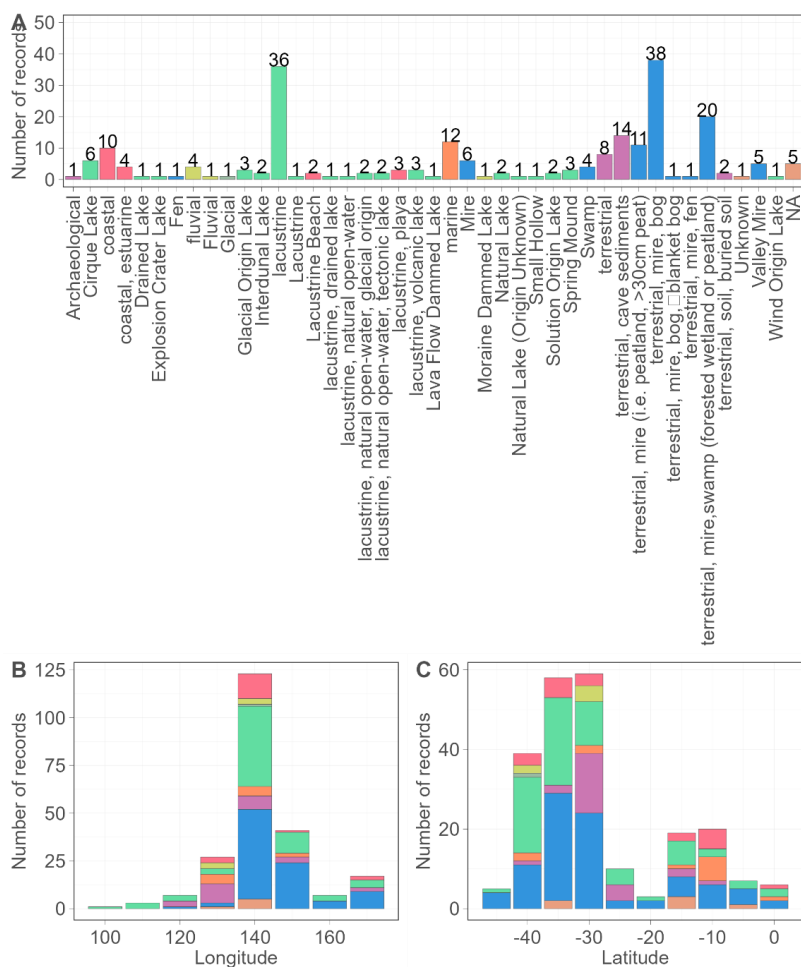



Figure A3, a) Number of records by sedimentary environment; b) Number of records by latitude, colours representing sedimentary environment from a; c) Number of records by

370   longitude, colours representing sedimentary environment from a.

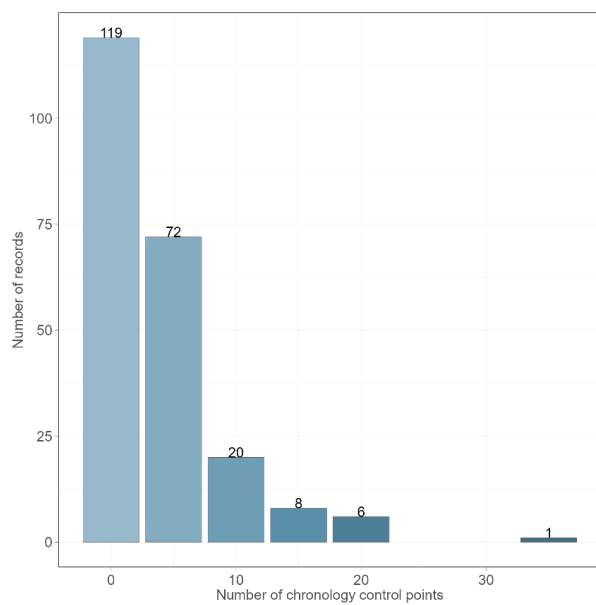



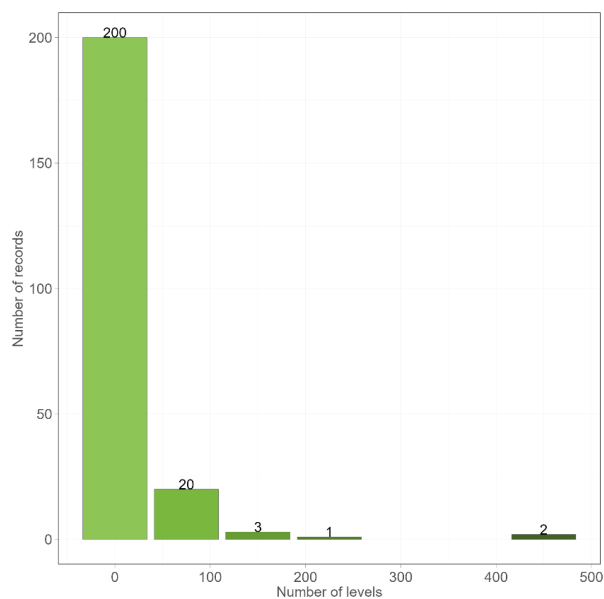
Figure A4, Number of records by number of chronological control points.

375

Figure A5, Number of records by number of levels.

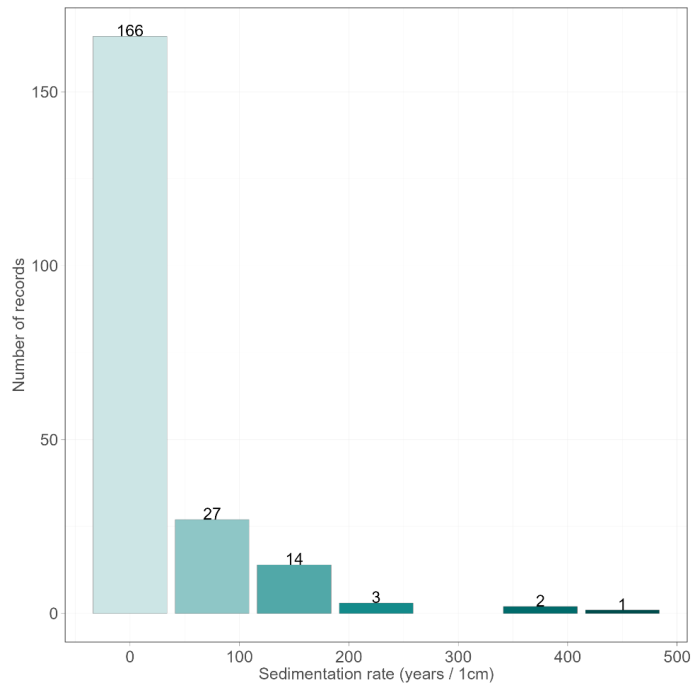



Figure A6, Number of records by sedimentation rate.