

APPENDIX 1: Introductory reading for macroecologists interested in learning more about Quaternary pollen analysis

Here we provide a selection of digitally available references that serve as introductory reading for macroecologists and multi-disciplinary readers. They are specifically aimed at sections within the *FOSSILPOL* Workflow related to:

- I. Extraction of sequences and depositional environments**
- II. Age-depth modelling**
- III. Pollen and spore taxonomy**
- IV. Taxonomic harmonisation and resolution**
- V. Quaternary pollen-analytical research in general**

It is not our intention to provide an exhaustive list of topics related to Quaternary palaeoecology, but only to target the key parts within the Workflow where these readings can help the user have the necessary background to start thinking about the criteria to apply. We recommend that users follow up the detailed references within these papers for more in-depth background reading.

I. Extraction of records and depositional environments

1) Extraction and processing of fossil pollen records

- a) 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., Nolan, C., Mauri, A., de Vernal, A., Izumi, K., Holmström, L., Marsicek, J., Goring, S., Sommer, P.S., Chaput, M., Kupriyanov, D., 2020. Pollen-based climate reconstruction techniques for late Quaternary studies. *Earth-Science Reviews* 210, 103384. <https://doi.org/10.1016/j.earscirev.2020.103384>
- b) Daniau, A.-L., Desprat, S., Aleman, J.C., Bremond, L., Davis, B., Fletcher, W., Marlon, J.R., Marquer, L., Montade, V., Morales-Molino, C., Naughton, F., Rius, D., Urrego, D.H., 2019. Terrestrial plant microfossils in palaeoenvironmental studies, pollen, microcharcoal and phytolith. Towards a comprehensive understanding of vegetation, fire and climate changes over the past one million years. *Revue de Micropaléontologie* 63, 1–35. <https://doi.org/10.1016/j.revmic.2019.02.001>

2) Understanding the relevance of pollen source areas:

- a) 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., Nolan, C., Mauri, A., de Vernal, A., Izumi, K., Holmström, L., Marsicek, J., Goring, S., Sommer, P.S., Chaput, M., Kupriyanov, D., 2020. Pollen-based climate reconstruction techniques for late Quaternary studies. *Earth-Science Reviews* 210, 103384. <https://doi.org/10.1016/j.earscirev.2020.103384>
- b) Daniau, A.-L., Desprat, S., Aleman, J.C., Bremond, L., Davis, B., Fletcher, W., Marlon, J.R., Marquer, L., Montade, V., Morales-Molino, C., Naughton, F., Rius, D., Urrego, D.H., 2019. Terrestrial plant microfossils in palaeoenvironmental studies, pollen, microcharcoal and phytolith. Towards a comprehensive understanding of vegetation, fire and climate changes over

- the past one million years. *Revue de Micropaléontologie* 63, 1–35.
<https://doi.org/10.1016/j.revmic.2019.02.001>
- c) Jacobson, G.L., Bradshaw, R.H.W., 1981. The selection of sites for paleovegetational studies. *Quaternary Research* 16, 80–96. [https://doi.org/10.1016/0033-5894\(81\)90129-0](https://doi.org/10.1016/0033-5894(81)90129-0)
 - d) Sugita, S., 1994. Pollen representation of vegetation in Quaternary sediments: Theory and method in patchy vegetation. *Journal of Ecology* 82, 881–897. <https://doi.org/10.2307/2261452>
- 3) Depositional environments and taphonomy**
- a) See Box in: Nieto-Lugilde, D., Blois, J.L., Bonet-García, F.J., Giesecke, T., Gil-Romera, G., Seddon, A., 2021. Time to better integrate paleoecological research infrastructures with neoecology to improve understanding of biodiversity long-term dynamics and to inform future conservation. *Environmental Research Letters* 16, 095005. <https://doi.org/10.1088/1748-9326/ac1b59>
 - b) Cleal, C., Pardoe, H.S., Berry, C.M., Cascales-Miñana, B., Davis, B.A.S., Diez, J.B., Filipova-Marinova, N.V. et al., 2021. Palaeobotanical experiences of plant diversity in deep time. 1: how well can we identify past plant diversity in the fossil record? *Palaeogeography, Palaeoclimatology, Palaeoecology* 576, 110481. <https://doi.org/10.1016/j.palaeo.2021.110481>
- 4) Dealing with depositional environments in multi-records analyses and databases:**
- a) Jacobson, G.L., Bradshaw, R.H.W., 1981. The selection of sites for paleovegetational studies. *Quaternary Research* 16, 80–96. [https://doi.org/10.1016/0033-5894\(81\)90129-0](https://doi.org/10.1016/0033-5894(81)90129-0)
 - b) Goring, S., Lacourse, T., Pellatt, M.G., Walker, I.R., Mathewes, R.W., 2010. Are pollen-based climate models improved by combining surface samples from soil and lacustrine substrates? *Review of Palaeobotany and Palynology* 162, 203–212. <https://doi.org/10.1016/j.revpalbo.2010.06.014>
 - c) Wilmshurst, J.M., McGlone, M.S., 2005. Origin of pollen and spores in surface lake sediments: Comparison of modern palynomorph assemblages in moss cushions, surface soils and surface lake sediments. *Review of Palaeobotany and Palynology* 136, 1–15. <https://doi.org/10.1016/j.revpalbo.2005.03.007>

II. Age-depth modelling

- 1) Blaauw, M. 2010. (Outdated) manual of Clam but with relevant explanations: <https://chrono.qub.ac.uk/blaauw/clam.html>
- 2) Blaauw, M., & Heegaard, E. (2012). Estimation of Age-Depth Relationships. In H. J. B. Birks, A. F. Lotter, S. Juggins, & J. P. Smol (Eds.), *Tracking Environmental Change Using Lake Sediments, Vol. 5: Data Handling and Numerical Techniques* (pp. 379–413). Springer Netherlands. https://doi.org/10.1007/978-94-007-2745-8_12
- 3) Grimm, E.C., Blaauw, M., Buck, C., Williams, J., 2014. Age models, chronologies, and databases workshop. *PAGES Magazine* 22, 104–104. <https://doi.org/10.22498/pages.22.2.104> and see <https://cp.copernicus.org/preprints/cp-2017-31/cp-2017-31-ED1.pdf>
- 4) Trachsel, M., Telford, R.J., 2017. All age–depth models are wrong, but are getting better. *The Holocene* 27, 860–869. <https://doi.org/10.1177/0959683616675939>

III. Pollen and spore taxonomy

Global online reference material for the identification of pollen:

- a) Hooghiemstra, H., van Geel, B., 1998. World list of Quaternary pollen and spore atlases. Review of Palaeobotany and Palynology 104, 157–182. [https://doi.org/10.1016/S0034-6667\(98\)00053-0](https://doi.org/10.1016/S0034-6667(98)00053-0)
- b) Weber, M., Ulrich, S., 2018. PalDat - A useful tool for identifying recent and fossil pollen. [Link](#)
- c) Global Pollen Project: [Link](#)
- d) Pollen and spores image database Göttingen: [Link](#)
- e) See also table 1 in Daniau et al. 2019. Revue de Micropaléontologie for a list of global pollen atlases.

IV. Taxonomic harmonisation and resolution

Existing harmonisation tables of fossil pollen data:

- a) Cao, X., Tian, F., Andreev, A., Anderson, P.M., Lozhkin, A.V., Bezrukova, E., Ni, J., Rudaya, N., Stobbe, A., Wiczorek, M., Herzschuh, U., 2020. A taxonomically harmonized and temporally standardized fossil pollen dataset from Siberia covering the last 40 kyr. Earth System Science Data 12, 119–135. <https://doi.org/10.5194/essd-12-119-2020>
- b) Gavin, D.G., Oswald, W.W., Wahl, E.R., Williams, J.W., 2003. A statistical approach to evaluating distance metrics and analog assignments for pollen records. Quaternary Research 60, 356–367. [https://doi.org/10.1016/S0033-5894\(03\)00088-7](https://doi.org/10.1016/S0033-5894(03)00088-7). Note: Called “P25” and includes 25 pollen taxa. Available in Neotoma package.
- c) Giesecke, T., Wolters, S., van Leeuwen, J.F.N., van der Knaap, P.W.O., Leydet, M., Brewer, S., 2019. Postglacial change of the floristic diversity gradient in Europe. Nature Communications 10, 5422. <https://doi.org/10.1038/s41467-019-13233-y>
- d) Herzschuh, U., Li, C., Böhmer, T., Postl, A.K., Heim, B., Andreev, A.A., Cao, X., Wiczorek, M., & Ni, J. 2022. LegacyPollen 1.0: a taxonomically harmonized global late Quaternary pollen dataset of 2831 records with standardized chronologies. Earth System Science Data 14: 3213–3227. <https://doi.org/10.5194/essd-14-3213-2022>
- e) Li, C., Postl, A.K., Böhmer, T., Cao, X., Dolman, A.M., Herzschuh, U., 2022. Harmonized chronologies of a global late Quaternary pollen dataset (LegacyAge 1.0). Earth System Science Data 14, 1331–1343. <https://doi.org/10.5194/essd-14-1331-2022>
- f) 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., Kuneš, P., Wolters, S., Seddon, A.W.R., Williams, J.W., 2021. Global acceleration in rates of vegetation change over the past 18,000 years. Science 372, 860–864. <https://doi.org/10.1126/science.abg1685>
- g) Phelps, L.N., Chevalier, M., Shanahan, T.M., Aleman, J.C., Courtney-Mustaphi, C., Kiahtipes, C.A., Broennimann, O., Marchant, R., Shekeine, J., Quick, L.J., Davis, B.A.S., Guisan, A., Manning, K., 2020. Asymmetric response of forest and grassy biomes to climate variability across the African Humid Period: influenced by anthropogenic disturbance? Ecography 43, 1118–1142. <https://doi.org/10.1111/ecog.04990>
- h) Sánchez Goñi, M.F., Desprat, S., Daniau, A.-L., Bassinot, F.C., Polanco-Martínez, J.M., Harrison, S.P., Allen, J.R.M., et al., 2005. Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications. Quaternary Science Reviews 24, 1828–1848. <https://doi.org/10.1016/j.quascirev.2005.03.005>. Note: “WhitmoreFull”, full list.

“WhitmoreSmall”, As WhitmoreFull but taxa for which both fully resolved and undifferentiated exist these taxa are summed. Both available in Neotoma package.

- i) Williams, J.W., Shuman, B., 2008. Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. *Quaternary Science Reviews* 27, 669–687. <https://doi.org/10.1016/j.quascirev.2008.01.004>.
Note: Called “WS64” and available in Neotoma package.

The influence of taxonomic resolution and harmonisation levels for different research purposes:

- a) Deza-Araujo, M., Morales-Molino, C., Conedera, M., Pezzatti, G.B., Pasta, S., Tinner, W., 2022. Influence of taxonomic resolution on the value of anthropogenic pollen indicators. *Vegetation History and Archaeobotany* 31, 67–84. <https://doi.org/10.1007/s00334-021-00838-x>
- b) Mitchell, E.A.D., Lamentowicz, M., Payne, R.J., Mazei, Y., 2014. Effect of taxonomic resolution on ecological and palaeoecological inference – a test using testate amoeba water table depth transfer functions. *Quaternary Science Reviews* 91, 62–69. <https://doi.org/10.1016/j.quascirev.2014.03.006>
- c) Rull, V. (2012) Palaeobiodiversity and taxonomic resolution: linking past trends with present patterns. *Journal of Biogeography* 39:1005–1006. <https://doi.org/10.1111/j.1365-2699.2012.02735.x>

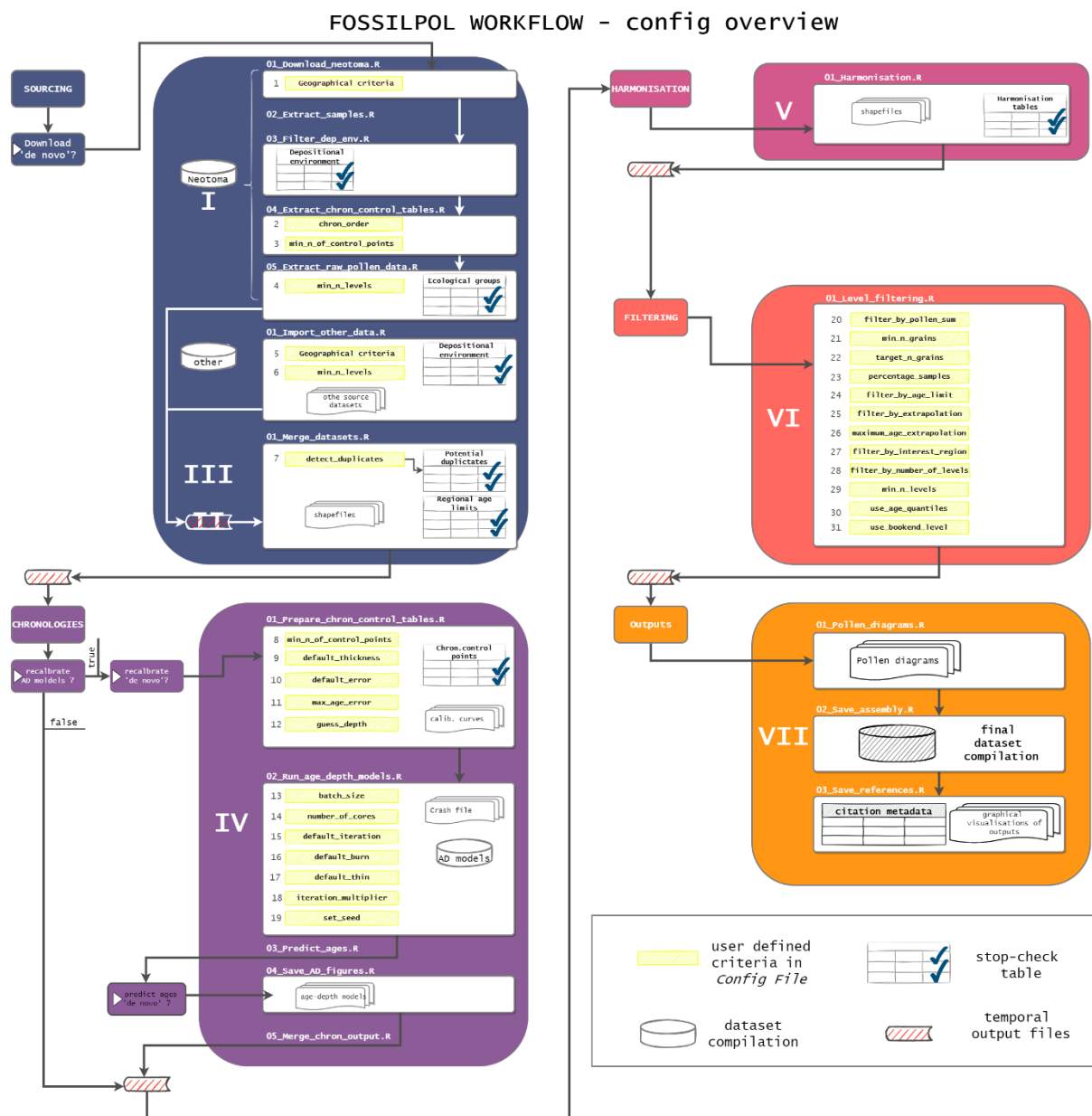
Difference between plant taxa and pollen types:

De Klerk, P., Joosten, H., 2007. The difference between pollen types and plant taxa: a plea for clarity and scientific freedom. *E & G Quaternary Science Journal* 56, 162–171. <https://doi.org/10.3285/eg.56.3.02>

V. Quaternary pollen-analytical research in general

- a) Birks, H.J.B., Berglund, B.E., 2018. One hundred years of Quaternary pollen analysis 1916–2016. *Vegetation History and Archaeobotany* 27, 271–309. <https://doi.org/10.1007/s00334-017-0630-2>
- b) Birks, H.J.B., 2019. Contributions of Quaternary botany to modern ecology and biogeography. *Plant Ecology & Diversity* 12, 189–385. <https://doi.org/10.1080/17550874.2019.1646831>
- c) Rull, V., 2020. Quaternary ecology, evolution, and biogeography, 1st ed. Elsevier. <https://www.elsevier.com/books/quaternary-ecology-evolution-and-biogeography/rull/978-0-12-820473-3>
- d) Dillon, E., Dunne, E., Womack, T., Kouvari, M., Larina, E., Claytor, J., . . . Zill, M. (2023). Challenges and directions in analytical paleobiology. *Paleobiology*, 1-17. <https://doi.org/10.1017/pab.2023.3>

165 APPENDIX 2: Detailed figure of the FOSSILPOL workflow with all steps included in the
 166 Config file.



168 APPENDIX 3: General template for FOSSILPOL Config file

```

169 #-----#
170 #
171 #
172 #           The FOSSILPOL workflow
173 #
174 #           Config file
175 #
176 #
177 #   O. Mottl, S. Flantua, K. Bhatta, V. Felde, A. Seddon
178 #                               2023
179 #
180 #-----#
181
182 # Configuration script with the variables that should be consistent
183 throughout
184 #   the whole repo. It loads packages, defines important variables,
185 #   authorises the user, and saves the config file.
186 # Points that require the user's attention are flagged by "[USER]" - flag,
187 # meaning that these are criteria that need to be checked by the user
188
189 # Version of the Workflow
190 workflow_version <-
191   "0.0.2"
192
193 # Set the current environment
194 current_env <- environment()
195
196 #-----#
197 # 1. Load packages -----
198 #-----#
199
200 if (
201   isFALSE(
202     exists("already_synch", envir = current_env)
203   )
204 ) {
205   already_synch <- FALSE
206 }
207
208 if (
209   isFALSE(already_synch)
210 ) {
211   library(here)
212   # Synchronise the package versions
213   renv::restore(
214     lockfile = here::here("renv/library_list.lock")
215   )
216   already_synch <- TRUE
217
218   # Save snapshot of package versions
219   # renv::snapshot(lockfile = "renv/library_list.lock") # do only for
220 update
221 }
222
223 # Define packages
224 package_list <-
225   c(
226     "RFossilpol",

```

```

227     "RUtilpol",
228     "here",
229     "tidyverse"
230   )
231
232   # Attach all packages
233   sapply(package_list, library, character.only = TRUE)
234
235   #-----#
236   # 2. Current date and working directory -----#
237   #-----#
238
239   current_date <- Sys.Date()
240
241   # Project directory is set up by the {here} package, Adjust if needed
242   current_dir <- here::here()
243
244   # Define the directory (external) for storing big data files
245   # Default is in the current project
246   data_storage_path <- current_dir # [USER]
247
248   # Create all essential folders
249   RFossilpol::util_make_datastorage_folders(
250     dir = data_storage_path # [config_criteria]
251   )
252
253   #-----#
254   # 3. Load functions -----#
255   #-----#
256
257   # Get a vector of general functions
258   fun_list <-
259     list.files(
260       path = "R/Functions/",
261       pattern = "*.R",
262       recursive = TRUE
263     )
264
265   # Load the function into the global environment
266   sapply(
267     paste0("R/Functions/", fun_list, sep = ""),
268     source
269   )
270
271   #-----#
272   # 4. Project dataset database -----#
273   #-----#
274
275   # Check the presence of a dataset database and create it if necessary
276   if (
277     isFALSE(
278       "project_dataset_database.rds" %in%
279       list.files(
280         paste0(
281           data_storage_path, # [config_criteria]
282           "/Data/Personal_database_storage"
283         )
284       )
285     )
286   )
287

```

```

288 ) {
289   project_dataset_database <-
290     RFossilpol::proj_db_class()
291
292   readr::write_rds(
293     project_dataset_database,
294     paste0(
295       data_storage_path, # [config_criteria]
296       "/Data/Personal_database_storage",
297       "/project_dataset_database.rds"
298     ),
299     compress = "gz"
300   )
301 }
302
303
304 #-----#
305 # 5. Define variables -----#
306 #-----#
307
308 #-----#
309 # 5.1. Define individual run -----#
310 #-----#
311
312 # Include/exclude Neotoma download in run
313 dataset_type <- "pollen"
314
315 # Selected variable element (proxy)
316 sel_var_element <- "pollen"
317
318 # Set geographical boundaries
319 long_min <- -180 # [USER]
320 long_max <- 180 # [USER]
321 lat_min <- -90 # [USER]
322 lat_max <- 90 # [USER]
323 alt_min <- NA # [USER]
324 alt_max <- NA # [USER]
325
326 neotoma_new_download <- TRUE
327
328 # Define access to datasets from other sources than Neotoma
329 use_other_datasource <- FALSE # [USER]
330 detect_duplicates <- TRUE # [USER]
331
332 # Include/exclude age modelling in run
333 recalib_AD_models <- TRUE # [USER]
334 calc_AD_models_denovo <- FALSE # [USER]
335 predict_ages_denovo <- FALSE # [USER]
336
337 # Select the final variables
338 select_final_variables <- TRUE # [USER]
339
340
341 #-----#
342 # 5.2. Chronology order -----#
343 #-----#
344
345 # Selected and preferred order of age type of existing chronologies in
346 Neotoma
347 chron_order <-
348   tibble::tibble(

```



```

349     order = seq(1, 6),
350     type = c(
351         "Varve years BP",
352         "Calibrated radiocarbon years BP",
353         "Calendar years BP",
354         "Radiocarbon years BP",
355         "Calendar years AD/BC",
356         NA
357     )
358 )
359
360
361 #-----#
362 # 5.5. Age depth models ----
363 #-----#
364
365 # Chronology needs to have at least X control points [example: X=2]
366 min_n_of_control_points <- 2 # [USER]
367
368 # If the thickness of a control point is missing, assign X cm [example:
369 X=1]
370 default_thickness <- 1 # [USER]
371
372 # If the age error of a control point is missing, assign X yr [example:
373 X=100]
374 default_error <- 100 # [USER]
375
376 # Maximum accepted age error of chron.control point [example: X=3000]
377 max_age_error <- 3000 # [USER]
378
379 # Depth at which "Guess" is accepted as a chronology control
380 # point [example: X=10]
381 guess_depth <- 10 # [USER]
382
383 # Bchron settings
384 number_of_cores <- parallel::detectCores() - 1
385 batch_size <- number_of_cores * 3
386 set_seed <- 1234
387
388 default_iteration <- 10e3
389 default_burn <- 2000
390 default_thin <- 8 # [USER]
391 iteration_multiplier <- 5 # [USER]
392
393
394 #-----#
395 # 5.6. Filtering criteria ----
396 #-----#
397
398 # Criteria to filter out stratigraphic levels and pollen records
399
400 #-----#
401
402 # Pollen sums
403 filter_by_pollen_sum <- TRUE # [USER]
404
405 # Each stratigraphic level of at least X individual pollen gains [example:
406 X=0]
407 min_n_grains <- 0 # [USER]
408 # Ideal number of counts
409 target_n_grains <- 100 # [USER]

```

```

410 # Threshold of number of samples with ideal counts
411 percentage_samples <- 0 # [USER]
412
413 #-----#
414
415 # Age limits
416 # Note that the actual ages have to be specified per region, defined
417 during
418 # the process of Workflow
419 filter_by_age_limit <- TRUE # [USER]
420
421 #-----#
422
423 # Maximum extrapolation
424 filter_by_extrapolation <- TRUE # [USER]
425
426 # How much age can be extrapolated beyond the oldest chronology control
427 point?
428 maximum_age_extrapolation <- Inf # [USER]
429
430 #-----#
431
432 # Beyond the period of interest
433 # Note that the actual ages have to be specified per region, defined
434 during
435 # the process of Workflow
436 filter_by_interest_region <- TRUE # [USER]
437
438 #-----#
439
440 # Number of stratigraphic levels
441 filter_by_number_of_levels <- TRUE # [USER]
442
443 # At least X number stratigraphic levels within the period
444 # of interest [example: X=3]
445 min_n_levels <- 3
446
447 #-----#
448
449 # Additional setting
450
451 # Should the 95th age quantile be used for data filtration?
452 # If FALSE (default), the estimated age will be used for all checks about
453 the
454 # age of a stratigraphic level. However, if TRUE, then the 95th age
455 quantile
456 # will be used.
457 # This will result in more stable data assembly between different results
458 # of age-depth modelling BUT require additional data preparation before
459 # the analytical part.
460 use_age_quantiles <- FALSE # [USER]
461
462 # Should all data filtration omit one additional stratigraphic level in
463 # the old period?
464 # If TRUE, all filtering will proceed normally but one additional "bookend"
465 # stratigraphic level will be always kept.
466 # A "bookend" stratigraphic level is a subsequential stratigraphic level
467 # older than the oldest stratigraphic level that passed the filtration.
468 # A "bookend" can help provide anchor information older than the period of
469 # interest.
470 use_bookend_level <- FALSE # [USER]

```

```

471
472
473 #-----#
474 # 6. Set graphical options -----
475 #-----#
476
477 # Set ggplot output
478 ggplot2::theme_set(
479   ggplot2::theme_classic()
480 )
481
482 # Define general
483 text_size <- 16 # [USER]
484 line_size <- 0.1 # [USER]
485 point_size <- 3 # [USER]
486
487 # Define output sizes
488 image_width <- 30 # [USER]
489 image_height <- 15 # [USER]
490 image_units <- "cm" # [USER]
491 image_dpi <- 300 # [USER]
492
493 # Define pallets
494
495 # Define common colours
496
497
498 #-----#
499 # 7. Save current config setting -----
500 #-----#
501
502 current_setting <-
503   RFossilpol::util_extract_config_data()
504

```

APPENDIX 4: FOSSILPOL Config file used for the Northern Europe example

```
507 The important changes in comparison to the default template are highlighted blue.
508 #-----#
509 #
510 #
511 #           Fossil pollen data from Scandinavia
512 #
513 #           Config file
514 #
515 #
516 #   O. Mottl, S. Flantua, K. Bhatta, V. Felde, A. Seddon
517 #               2023
518 #
519 #-----#
520
521 # Configuration script with the variables that should be consistent
522 throughout
523 #   the whole repo. It loads packages, defines important variables,
524 #   authorises the user, and saves the config file.
525 # Points that require the user's attention are flagged by "[USER]" - flag,
526 # meaning that these are criteria that need to be checked by the user
527
528 # Version of the Workflow
529 workflow_version <-
530   "0.0.2"
531
532 # Set the current environment
533 current_env <- environment()
534
535 #-----#
536 # 1. Load packages ----
537 #-----#
538
539 if (
540   isFALSE(
541     exists("already_synch", envir = current_env)
542   )
543 ) {
544   already_synch <- FALSE
545 }
546
547 if (
548   isFALSE(already_synch)
549 ) {
550   library(here)
551   # Synchronise the package versions
552   renv::restore(
553     lockfile = here::here("renv/library_list.lock")
554   )
555   already_synch <- TRUE
556
557   # Save snapshot of package versions
558   # renv::snapshot(lockfile = "renv/library_list.lock") # do only for
559   update
560 }
561
```

```

562 # Define packages
563 package_list <-
564   c(
565     "RFossilpol",
566     "RUtilpol",
567     "here",
568     "tidyverse"
569   )
570
571 # Attach all packages
572 sapply(package_list, library, character.only = TRUE)
573
574 #-----#
575 # 2. Current date and working directory -----#
576 #-----#
577
578 current_date <- Sys.Date()
579
580 # Project directory is set up by the {here} package, Adjust if needed
581 current_dir <- here::here()
582
583 # Define the directory (external) for storing big data files
584 # Default is in the current project
585 data_storage_path <- current_dir # [USER]
586
587 # Create all essential folders
588 RFossilpol::util_make_datastorage_folders(
589   dir = data_storage_path # [config_criteria]
590 )
591
592 #-----#
593 # 3. Load functions -----#
594 #-----#
595
596 # Get a vector of general functions
597 fun_list <-
598   list.files(
599     path = "R/Functions/",
600     pattern = "*.R",
601     recursive = TRUE
602   )
603
604 # Load the function into the global environment
605 sapply(
606   paste0("R/Functions/", fun_list, sep = ""),
607   source
608 )
609
610 #-----#
611 # 4. Project dataset database -----#
612 #-----#
613
614 # Check the presence of a dataset database and create it if necessary
615 if (
616   isFALSE(
617     "project_dataset_database.rds" %in%
618     list.files(
619       paste0(
620         data_storage_path, # [config_criteria]

```

```

623         "/Data/Personal_database_storage"
624     )
625 )
626 )
627 ) {
628     project_dataset_database <-
629         RFossilpol::proj_db_class()
630
631     readr::write_rds(
632         project_dataset_database,
633         paste0(
634             data_storage_path, # [config_criteria]
635             "/Data/Personal_database_storage",
636             "/project_dataset_database.rds"
637         ),
638         compress = "gz"
639     )
640 }
641
642
643 #-----#
644 # 5. Define variables -----#
645 #-----#
646
647 #-----#
648 # 5.1. Define individual run -----#
649 #-----#
650
651 # Include/exclude Neotoma download in run
652 dataset_type <- "pollen"
653
654 # Selected variable element (proxy)
655 sel_var_element <- "pollen"
656
657 # Set geographical boundaries
658 long_min <- 5 # [USER]
659 long_max <- 30 # [USER]
660 lat_min <- 55 # [USER]
661 lat_max <- 73 # [USER]
662 alt_min <- NA # [USER]
663 alt_max <- NA # [USER]
664
665 neotoma_new_download <- TRUE
666
667 # Define access to datasets from other sources than Neotoma
668 use_other_datasource <- FALSE # [USER]
669 detect_duplicates <- TRUE # [USER]
670
671 # Include/exclude age modelling in run
672 recalib_AD_models <- TRUE # [USER]
673 calc_AD_models_denovo <- FALSE # [USER]
674 predict_ages_denovo <- FALSE # [USER]
675
676 # Select the final variables
677 select_final_variables <- TRUE # [USER]
678
679
680 #-----#
681 # 5.2. Chronology order -----#
682 #-----#
683

```

```

684 # Selected and preferred order of age type of existing chronologies in
685 Neotoma
686 chron_order <-
687   tibble::tibble(
688     order = seq(1, 6),
689     type = c(
690       "Varve years BP",
691       "Calibrated radiocarbon years BP",
692       "Calendar years BP",
693       "Radiocarbon years BP",
694       "Calendar years AD/BC",
695       NA
696     )
697   )
698
699
700 #-----#
701 # 5.5. Age depth models ----
702 #-----#
703
704 # Chronology needs to have at least X control points [example: X=2]
705 min_n_of_control_points <- 5 # [USER]
706
707 # If the thickness of a control point is missing, assign X cm [example:
708 X=1]
709 default_thickness <- 1 # [USER]
710
711 # If the age error of a control point is missing, assign X yr [example:
712 X=100]
713 default_error <- 100 # [USER]
714
715 # Maximum accepted age error of chron.control point [example: X=3000]
716 max_age_error <- 3000 # [USER]
717
718 # Depth at which "Guess" is accepted as a chronology control
719 # point [example: X=10]
720 guess_depth <- 10 # [USER]
721
722 # Bchron settings
723 number_of_cores <- parallel::detectCores() - 1
724 batch_size <- number_of_cores * 3
725 set_seed <- 1234
726
727 default_iteration <- 10e3
728 default_burn <- 2000
729 default_thin <- 8 # [USER]
730 iteration_multiplier <- 5 # [USER]
731
732 #-----#
733 # 5.6. Filtering criteria ----
734 #-----#
735
736 # Criteria to filter out stratigraphic levels and pollen records
737
738 #-----#
739
740 # Pollen sums
741 filter_by_pollen_sum <- TRUE # [USER]
742
743

```

```

744 # Each stratigraphic level of at least X individual pollen gains [example:
745 X=0]
746 min_n_grains <- 100 # [USER]
747 # Ideal number of counts
748 target_n_grains <- 150 # [USER]
749 # Threshold of number of samples with ideal counts
750 percentage_samples <- 75 # [USER]
751
752 #-----#
753
754 # Age limits
755 # Note that the actual ages have to be specified per region, defined
756 during
757 # the process of Workflow
758 filter_by_age_limit <- TRUE # [USER]
759
760 #-----#
761
762 # Maximum extrapolation
763 filter_by_extrapolation <- TRUE # [USER]
764
765 # How much age can be extrapolated beyond the oldest chronology control
766 point?
767 maximum_age_extrapolation <- 3000 # [USER]
768
769 #-----#
770
771 # Beyond the period of interest
772 # Note that the actual ages have to be specified per region, defined
773 during
774 # the process of Workflow
775 filter_by_interest_region <- TRUE # [USER]
776
777 #-----#
778
779 # Number of stratigraphic levels
780 filter_by_number_of_levels <- TRUE # [USER]
781
782 # At least X number of stratigraphic levels within the period
783 # of interest [example: X=3]
784 min_n_levels <- 5
785
786 #-----#
787
788 # Additional setting
789
790 # Should the 95th age quantile be used for data filtration?
791 # If FALSE (default), the estimated age will be used for all checks about
792 the
793 # age of a stratigraphic level. However, if TRUE, then the 95th age
794 quantile
795 # will be used.
796 # This will result in more stable data assembly between different results
797 # of age-depth modelling BUT require additional data preparation before
798 # the analytical part.
799 use_age_quantiles <- FALSE # [USER]
800
801 # Should all data filtration omit one additional stratigraphic level in
802 # the old period?
803 # If TRUE, all filtering will proceed normally but one additional "bookend"
804 # stratigraphic level will be always kept.

```



```

805 # A "bookend" stratigraphic level is a subsequential stratigraphic level
806 #   older than the oldest stratigraphic level that passed the filtration.
807 # A "bookend" can help provide anchor information older than the period of
808 #   interest.
809 use_bookend_level <- FALSE # [USER]
810
811
812 #-----#
813 # 6. Set graphical options -----
814 #-----#
815
816 # Set ggplot output
817 ggplot2::theme_set(
818   ggplot2::theme_classic()
819 )
820
821 # Define general
822 text_size <- 16 # [USER]
823 line_size <- 0.1 # [USER]
824 point_size <- 3 # [USER]
825
826 # Define output sizes
827 image_width <- 30 # [USER]
828 image_height <- 15 # [USER]
829 image_units <- "cm" # [USER]
830 image_dpi <- 300 # [USER]
831
832
833 # Define common colours
834 col_gray_dark <- "gray30"
835 col_gray_middle <- "gray50"
836 col_gray_light <- "gray90"
837
838 ## Main colour
839 col_orange_dark <- "#E17D00"
840 col_orange_light <- "#ff9715"
841
842 ## Complementary
843 col_compl_blue <- "#00507B"
844
845 ## Analogous
846 col_ana_red <- "#FF2215"
847
848 # Define pallets
849 palette_generic <-
850   c(
851     "#FF9715",
852     "#FF685F",
853     "#D3588B",
854     "#8C5C99",
855     "#4A5783"
856   )
857
858 palette_matching <-
859   c(
860     "#00987E",
861     "#00D1B2",
862     "#B9A89B",
863     "#524439",
864     "#FF9715"
865   )

```

```

866
867 palette_shades <-
868   c(
869     "#570200",
870     "#7B2B00",
871     "#A54E00",
872     "#D17200",
873     "#FF9715"
874   )
875
876
877 #-----#
878 # 7. Save current config setting -----
879 #-----#
880
881 current_setting <-
882   RFossilpol::util_extract_config_data()
883

```