Database interoperability, uncertainty quantification and reproducible workflows in the paleogeosciences

Nicholas McKay

Northern Arizona University
(mailto:#)nick@nau.edu (mailto:nick@nau.edu)

Julien Emile-Geay

University of Southern California
(mailto:#)julieneg@usc.edu (mailto:julieneg@usc.edu)

Deborah Khider

University of Southern California
(mailto:#)Khider@usc.edu (mailto:Khider@usc.edu)

2022/04/22

1 Abstract

The paleogeosciences are becoming more and more interdisciplinary, and studies increasingly rely on large collections of data derived from multiple data repositories. Integrating diverse datasets from multiple sources into complex workflows increases the challenge of creating reproducible and open science, as data formats and tools are often noninteroperable, requiring manual manipulation of data into standardized formats, resulting in a disconnect in data provenance and confounding reproducibility. Here we present a notebook that demonstrates

how the Linked PaleoData (LiPD) framework is used as an interchange format to allow data from multiple data sources to be integrated in a complex workflow using emerging packages in R for geochronological uncertainty quantification and abrupt change detection. Specifically, in this notebook, we use the neotoma2 and lipdR packages to access paleoecological data from the Neotoma Database, and paleoclimate data from compilations hosted on Lipdverse. Age uncertainties for datasets from both sources are then quantified using the geoChronR package, and those data, and their associated age uncertainties, are then investigated for abrupt changes using the actR package, with accompanying visualizations. The result is an integrated, reproducible workflow in R that demonstrates how this complex series of multisource data integration, analysis and visualization can be integrated into an efficient, open scientific narrative.

2 Purpose

This notebook showcases a workflow that uses emerging tools to efficiently access a pollen dataset from Neotoma (Williams et al. 2018), create an age model using Bacon (Blaauw and Christen 2011), store the age ensemble and create visualizations using `geoChronR` (N. P. McKay, Emile-Geay, and Khider 2021), before finally conducting age-uncertain change detection using the Abrupt Change Toolkit in R (actR) (N. McKay and Emile-Geay 2022).

3 Technical contributions

This notebook highlights recent advances in multiple R packages

- The new `neotoma2` package (Dominguez Vidana and Goring 2022)
- An overhauled `neotoma2lipd()` function in the `lipdR` package (Heiser and McKay 2022)
- Using lipd objects and `geoChronR` (N. P. McKay, Emile-Geay, and Khider 2021) to create age models and store ensemble data
- Conducting robust, age-uncertain, abrupt change analysis on neotoma data using `actR` (N. McKay and Emile-Geay 2022)

4 Methodology

`neotoma2` leverages Neotoma's API v2.0 (Goring 2022). `lipdR::neotoma2lipd()` converts `neotoma2` site objects into `lipdR` lipd objects. This notebook uses Bacon Bacon (Blaauw and Christen 2011) to create an ensemble age-depth model within the `geoChronR` package (N. P. McKay, Emile-Geay, and Khider 2021). The changepoint detection algorithm implemented in `actR` (N. McKay and Emile-Geay 2022) which is used in this notebook relies heavily on the `changepoint` package Killick, Haynes, and Eckley (2016).

5 Results

This notebook demonstrates how the Linked PaleoData (LiPD) framework can be used as an interchange format to allow data from multiple data sources to be integrated in a complex workflow using emerging packages in R for geochronological uncertainty quantification and abrupt change detection. Efficient workflows that support data from multiple sources are an important component of reproducible science.

6 Funding

This notebook describes several LinkedEarth activities. LinkedEarth is a community of paleoscientists working to develop standards and software to enable paleoscience in the era of Big Data. This community produces data products and standards, software, cyberinfrastructure, and training opportunities. LinkedEarth was launched as part of an EarthCube Integrative Activities project funded project (ICER-1540996), and is currently supported by an EarthCube Data Capabilities project (ICER-2126510).

Neotoma-LiPD interoperability, and the Abrubt Change Toolkit in R (actR) are LinkedEarth activities that are funded as part of the Belmont Forum's Science-driven e-Infrastructure Innovation (SEI) initiative. In the US, this is NSF award ICER-1929460.

7 Citation

Include the recommended citation for the document. You may choose to use Zenodo, or another service (such as figShare or your University services) to obtain a DOI for your code repository. If you are using GitHub, consider adding a GitHub Citation file (https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-citation-files). Ensure that the title in the suggested citation, and the authors match those in your YAML header.

For example:

- McKay, N., Emile Geay, J., Khider, D, 2022. Database interoperability, uncertainty quantification and reproducible workflows in the paleogeosciences. Accessed 4/22/2022 at https://github.com/nickmckay/EC22-neotoma-actR (https://github.com/nickmckay/EC22-neotoma-actR)

8 Acknowledgements

Thank you to Socorro Dominguez Vidana and Simon Goring for their work on the `neotoma2` and to the scientists who generated and shared the Bambili pollen data and the MD03-2707 paleotemperature data.

9 Setup

You will need to download the Rmd, HTML and bibtex file from the earthcube repository, store it in a single directory and then set up a project (see this link (https://swcarpentry.github.io/r-novice-gapminder/02-project-intro/) for further discussion of this.) Then you can either knit the markdown file, or run the cells interactively.

9.1 Packages

Install all the required packages in the R setup section at the head of this document. As noted previously, use pacman to do the installation. Alternatively, use the `install.R` script in the repository to help install some of the github packages prior to the pacman cells. Running the notebook on mybinder (https://mybinder.org/v2/gh/nickmckay/EC22-neotoma-actR/main?urlpath=rstudio), is also an option.

9.2 Data processing and analysis

The paleogeosciences are becoming more and more interdisciplinary, and studies increasingly rely on large collections of data derived from multiple data repositories. Integrating diverse datasets from multiple sources into complex workflows increases the challenge of creating reproducible and open science, as data formats and tools are often noninteroperable, requiring manual manipulation of data into standardized formats, resulting in a

disconnect in data provenance and confounding reproducibility. Here, we illustrate a use case where a scientist wants to conduct age-uncertain abrupt-change analysis on two datasets from Africa - a pollen record from Lake Bambili, and a temperature reconstruction from the Gulf of Guinea.

9.2.1 Lake Bambili

Lake Bambili is a high-elevation (2273 masl) crater lake in Cameroon (05° 56′ 11.9 N, 10°14′ 31.6 E) in the highlands of equatorial West Africa. Recently, (Lézine et al. 2019) published a study using pollen data to show changes in forest composition over the past 90,000 years. In this notebook, we'll take a look at abrupt shifts in these pollen assemblages, and compare them changes in a nearby climate record from the Gulf of Guinea.

Thankfully, Lezine et al. (2019) archived their pollen assemblage data at Neotoma, and we'll use the `neotoma2` package (Dominguez Vidana and Goring 2022) to access the data, in this case, searching by sitename.

```
L <- neotoma2::get_sites(sitename = "Bambili 2") %>% #Find the site of interest
  neotoma2::get_downloads() %>% #download the data for this site
  lipdR::neotoma2lipd() #convert the site object into a LiPD object
```

Now that we've downloaded the data and converted it to a LiPD object, it's ready to be used in LinkedEarth's ecosystem of tools built around the LiPD standard (e.g., geoChronR, actR, pyleoclim).

First, let's quickly create a map using the `mapLipd()` function in geoChronR to visualize Lake Bambili's location in equatorial west Africa.

```
mapLipd(L,map.type = "stamen",extend.range = 5)
```

Figure 1. Map of Africa showing the location of Lake Bambili.

9.2.2 Age modelling

In this notebook, we want to take a look at some of the abrupt changes in ecosystems recorded by pollen assemblages in this lake, while propagating age uncertainties. Neotoma does not store age ensembles, so we will need to generate a new age model. Here we'll use Bacon (Blaauw and Christen 2011), one of the methods built into geoChronR, to create an age model and import the ensembles into the LiPD structure. Here, we've specified alll of the parameters so it will run without input, but it's often convenient to call `runBacon()` with fewer inputs and run it interactively.

```
L <- geoChronR::runBacon(L,
              lab.id.var = 'neotomaChronConrolId',
              age.14c.var = NULL,
              age.14c.uncertainty.var = NULL,
              age.var = 'age',
              age.uncertainty.var = 'ageUnc',
              depth.var = 'depth',
              reservoir.age.14c.var = NULL,
              reservoir.age.14c.uncertainty.var = NULL,
              rejected.ages.var = NULL,
              plot.pdf = FALSE,
              accept.suggestions = TRUE)
```

Now that the age model has finished running and the results are stored in the LiPD object, let's plot the results.

```
plotChronEns(L) + ggtitle("Age - depth model for Lake Bambili")
```

```
## [1] "Found it! Moving on..."
## [1] "Found it! Moving on..."
## [1] "plotting your chron ensemble. This make take a few seconds..."
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```

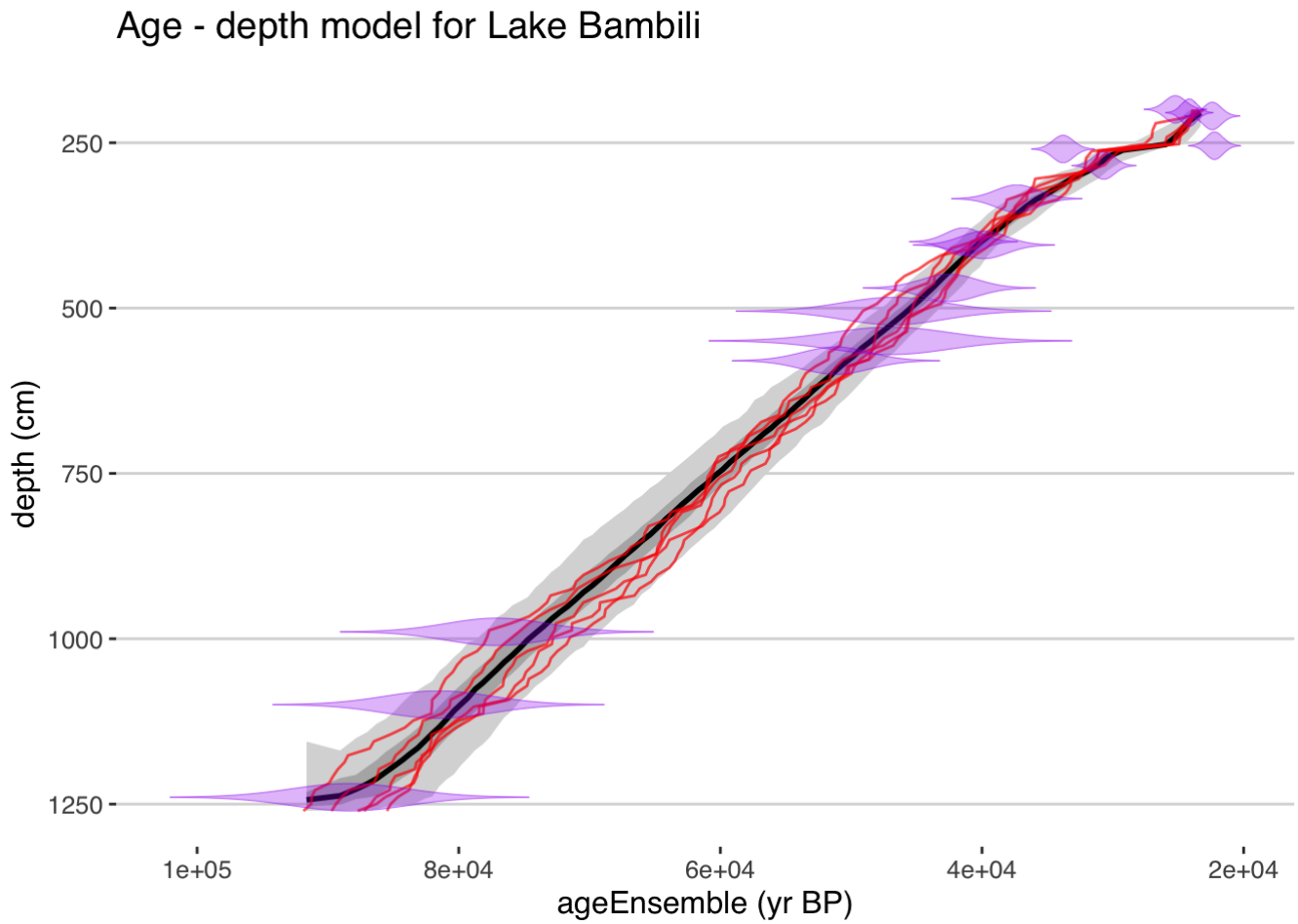

Figure 2. Age depth model for Lake Bambili, generated using Bacon. Ages (and their uncertainties) are shown as purple distributions. The median age model is shown in black, along with the 50 and 95% highest probability density regions in dark and light gray, respectively. Five randomly selected age ensemble members are shown in red.

Great, the uncertainties represented here are what we'd like to propagate through the abrupt change analysis. To do this, the first step is to "map" the age ensemble from the age model to the depths of the pollen data stored in the paleoData tables.

```
L <- mapAgeEnsembleToPaleoData(L)
```

```
## [1] "Bambili2.Lzine.2019"
## [1] "Looking for age ensemble...."
## [1] "Found it! Moving on..."
## [1] "Found it! Moving on..."
## [1] "getting depth from the paleodata table..."
## [1] "Found it! Moving on..."
```

Now that that's done, let's select the paleoData age ensemble for future reference.

```
ageEnsemble <- selectData(L,"ageEnsemble")
```

```
## [1] "Found it! Moving on..."
```

9.2.3 Paleoecological data analysis

Now let's do some analysis on the paleoecological data. For this example, we will just perform a simple calculation. We're interested in percent of pollen counts that are classified as trees or shrubs by their "ecological group."

To do this, we will take advantage of lipdR's integration with tidyverse tools in R. So our first step will be to convert the LiPD dataset into a specialized "tibble" object. It's a regular tibble, where each row is a variable in the dataset, which includes all the metadata, and nested variables for time and paleodata values. We then use `dplyr` tools to calculate the total amount of tree and shrub pollen counts for each sample. Finally, we use the `magrittr` pipe for efficiency and clarity. See the comments in the code block for step by step explanations.

```
TRSH <- as.lipdTsTibble(L) %>% #convert the LiPD object to a LiPD Timeseries Tibble obje
ct
  dplyr::filter(paleoData_ecologicalgroup == "TRSH") %>% #retain only the variables that
are in the trees and shrubs ecological group
  lipdR::tabulateTs(time.var = "age") %>% #convert this filtered lipd-ts-tibble to a tra
ditional paleoecological data table, with observations in the rows and age and pollen va
riables in the columns
  dplyr::rowwise() %>% #set up the analysis to row runwise
  dplyr::mutate(total = sum(c_across(-age),na.rm = TRUE)) #and use mutate to calculate a
new column, the sum of all pollen counts.
```

Great, we now have a regular tibble that includes all our sample levels, and a new "total" column that we've calculated. This will be the numerator in our "percent trees and shrubs" calculation. Let's repeat this process but with all of the non-aquatic pollen groups to ge the denominator.

```
total <- as.lipdTsTibble(L) %>%
  dplyr::filter(paleoData_ecologicalgroup %in% c("TRSH","UPHE","VACR","SUCC","MANG")) %
>%
  tabulateTs(time.var = "age") %>%
  rowwise() %>%
  mutate(total = sum(c_across(-age),na.rm = TRUE))
```

Now we can calculate the percentage of tree and shrub pollen relative to total non-aquatic pollen.

```
treeShrubPercent <- TRSH$total/total$total * 100
```

Now let's plot that percentage, along with the age ensemble, to visualize these data and their corresponding age uncertainties.

```
geoChronR::plotTimeseriesEnsRibbons(X = ageEnsemble, Y =  treeShrubPercent) + ylab("Rela
tive abundance of tree and shrub pollen (%)") + ggtitle("Tree/shrub pollen abundance at
 Lake Bambili.")
```

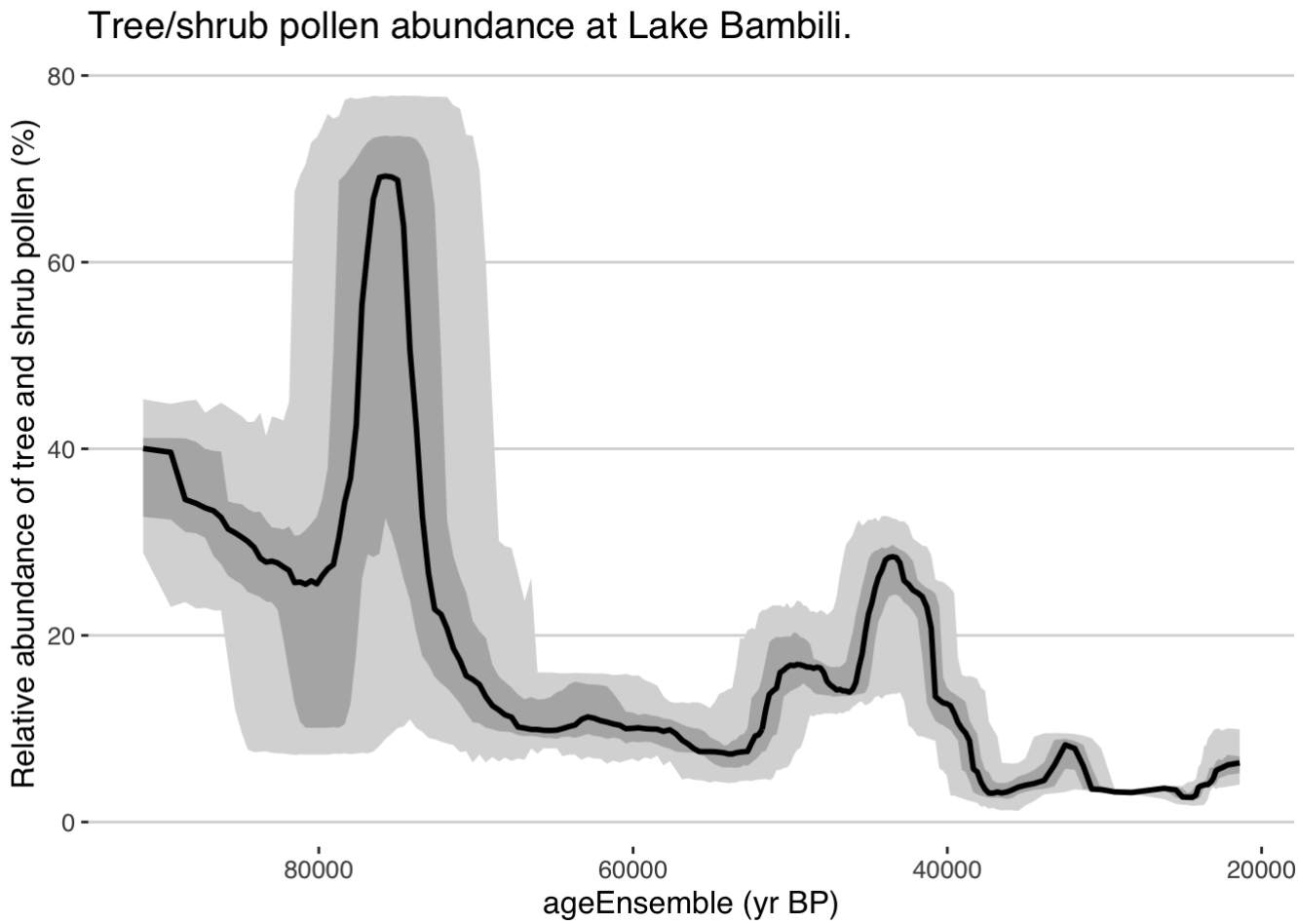

Figure 3. The relative abundance of tree and shrub pollen at Lake Bambili from ca. 90 to 20 ka, with associated age uncertainties. The median estimate is shown in black, along with the 50 and 95% highest probability density regions in dark and light gray, respectively.

Interesting, it does look like there are some rapid changes in this dataset that are worth further exploration.

9.2.4 Abrupt changes equatorial west African forest composition

We've now used the data we accessed from neotoma to generate an age ensemble, and a derived paleoecological indicator, and we're ready to analyze these data to test whether or not the apparent abrupt shifts stand out given age uncertainties and a robust null hypothesis. To do this, we'll use the "abrupt change toolkit in R (actR)" package to test for shifts in the mean of these data. actR can accept lipd-ts-tibble data directly, or data, ensembles and metadata separately. In this case, since we generated the paleoecological indicator outside of LiPD, we'll use the latter strategy and specify the data explicitly.

```
treeShrubMeanShift <- actR::detectShift(time = ageEnsemble$values, #matrix of age ensemb
le data
                  time.variable.name = "age", #variable name
                  time.units = "BP", # and units
                  vals = treeShrubPercent, #our paleoecological indicator (This could al
so be an ensemble if desired)
                  vals.variable.name = "Tree/Shrub", #variable name
                  vals.units = "%", # and units
                  summary.bin.step = 500, #over what timestep should we examine the resu
lts?
                  time.range = c(20000,90000)) #over what time range should we conduct t
he analysis.
```

```
## Warning in if (!is.na(time.range)) {: the condition has length > 1 and only the
## first element will be used
```

```
## Testing null hypothesis with 100 simulations, each with 100 ensemble members.
## This will probably take about 2 minutes
```

Because this methodology tests for abrupt shifts across the age ensemble and repeats the entire process with a null model to evaluate the robust null hypothesis level, it can take awhile to run.

Let's take a look at the results. The `print()` or `summarize()` methods will give some key details.

```
print(treeShrubMeanShift)
```

```
## Tree/Shrub: Shift in Mean results
## Searched for Shift in Mean with a minimum segment length of 1 years, summarizing the
results over windows of 500 years.
##
## Overall result: Detected 2 Shift in Mean event(s) that were significant at the alpha
< 0.05 level
## # A tibble: 2 × 3
##   time_start time_end empirical_pvalue
##        <dbl>    <dbl>            <dbl>
## 1     70153.   70653.           0.0200
## 2     65653.   66153.           0.0400
## Time uncertainty considered? TRUE Time ensemble supplied (n = 1000)
## Paleo uncertainty considered? TRUE Values ensemble generated in `propagateUncertainti
es()`)
## Error propagation ensemble members = 100
## Null hypothesis testing ensemble members = 100
## Members simulated using isospectral method
##
## Parameter choices:
## cpt.fun = changepoint::cpt.mean
## minimum.segment.length = 1
## method = AMOC
## penalty = MBIC
## ncpts.max = 1
```

These summaries are helpful, but ultimately, a visualization is often much more useful. Let's plot the results:

```
treeShrubPlot <- plot(treeShrubMeanShift)
```

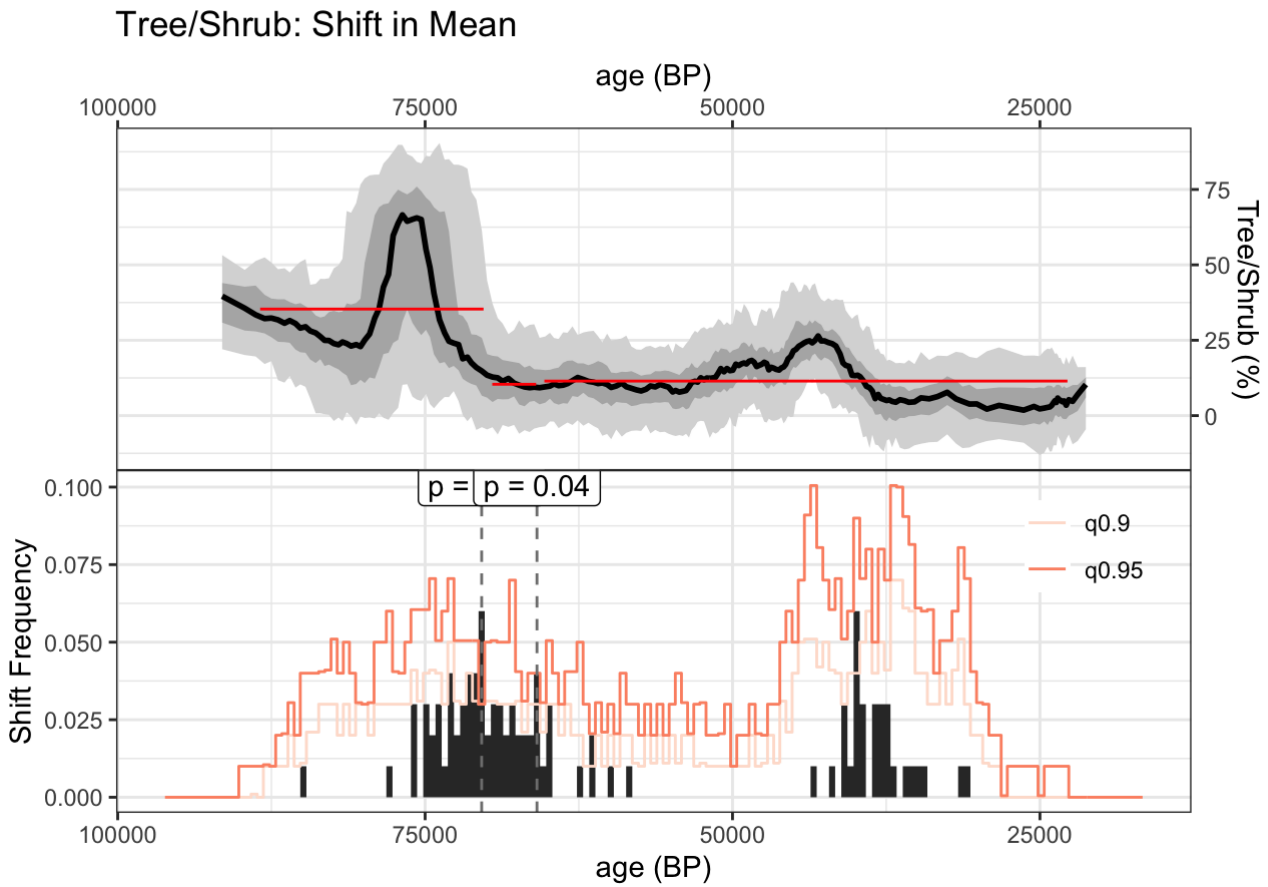

Figure 4. Results of the actR shift detection analysis for Lake Bambili tree and shrub abundance. Top panel: as in figure 3, except with red lines identifying the mean of intervals separated by significant detected shifts in the mean. Bottom panel: The black bars show the fraction of ensemble members for which a shift was detected during a each 500 year window, and the orange lines show 90 and 95th percentile thresholds results of the robust null hypothesis testing. Where the observed frequencies exceed the null hypothesis marks intervals where the detected shifts are robust at that confidence level.

Figure 4 visualizes many of the features of this analysis. First we see the identified intervals of abrupt change, as changes in the mean represented by the red line on the top panel, and as the black bars in the bottom panel, which show the proportion of ensemble members with detected abrupt changes during each 500-year window. Comparing this to the 95th percentile of changes found in the null model makes it clear that the shift that occurred about 71,000 years ago is a robust interval of abrupt change.

9.2.5 Links between climate and ecological change

Now that we've identified an interval of rapid change, let's take a look at a nearby climate record, a sea surface temperature (SST) reconstruction from the Gulf of Guinea (Weldeab et al. 2007), to take a first look of whether the change in ecological communities might be driven by climate. Neotoma doesn't host SST reconstuctions, so we'll get these data from lipdverse.org a community repository of LiPD datasets.

```
MD03_2707 <- readLipd("https://lipdverse.org/Temp12k/1_0_2/MD03_2707.Weldeab.2007.ensembles.lpd")
```

```
## [1] "Loading 1 datasets from /Volumes/data/tempdir//RtmpTqmA1m/MD03_2707.Weldeab.200
7.ensembles.lpd..."
## [1] "reading: MD03_2707.Weldeab.2007.ensembles.lpd"
```

And as before, let's take a look at the map.

```
mapLipd(MD03_2707,map.type = "stamen",extend.range = 5)
```

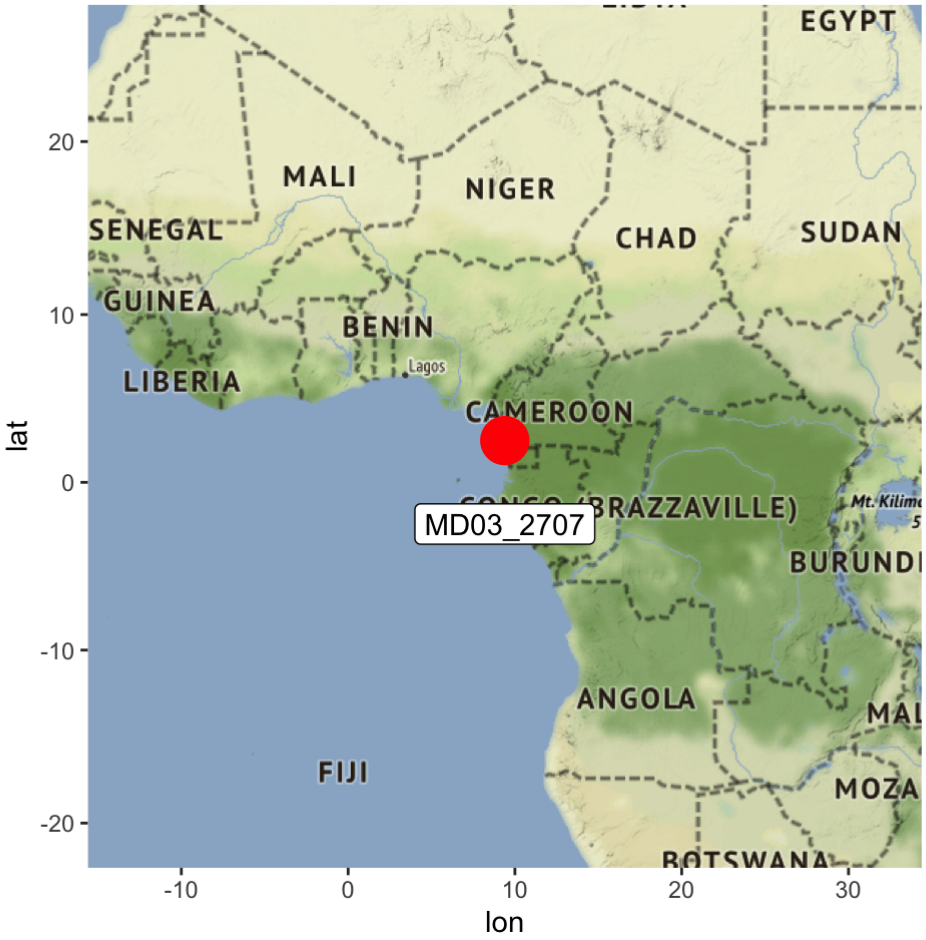

Figure 5. Location of core MD03-2707 in the Gulf of Guinea.

This dataset, originally published by Weldeab et al. (2007) and included in the Temperature 12k compilation (Kaufman et al. 2020), already includes age and temperature ensembles, so we don't have to run another age model. We still will need to map the ensembles to the paleodata and convert it to a lipd-ts-tibble though.

```
MD03_2707_ts <- mapAgeEnsembleToPaleoData(MD03_2707) %>% as.lipdTsTibble()
```

```
## [1] "MD03_2707.Weldeab.2007"
## [1] "Looking for age ensemble...."
## [1] "Found it! Moving on..."
## [1] "Found it! Moving on..."
## [1] "getting depth from the paleodata table..."
## [1] "Found it! Moving on..."
```

As before, now we will use `actR` to detect shifts in the mean, while propagating uncertainties, and comparing the results to a robust null hypothesis. This time, since all the data are stored in the LiPD object, we can just pass in the lipd-ts-tibble object directly, and then tell it which dataset we'd like to analyze.

```
MD03_2707_ms <-  actR::detectShift(MD03_2707_ts, #the lipd-ts-tibble that includes all t
he data
                   vals.variable.name = "SST_from_d18o_ruber_pink_ensemble", #the variab
le to analyze
                  summary.bin.step = 500, #same as before
                  time.range = c(20000,90000)) #same as before
```

```
## Selected SST_from_d18o_ruber_pink_ensemble
## timeUnits is at least partially absent in the input data
## Testing null hypothesis with 100 simulations, each with 100 ensemble members.
## This will probably take about 1 minutes
```

And as before, we can print out the summary:

```
MD03_2707_ms
```

```
## MD03_2707.Weldeab.2007 - SST_from_d18o_ruber_pink_ensemble: Shift in Mean results
## Searched for Shift in Mean with a minimum segment length of 1 years, summarizing the
results over windows of 500 years.
##
## Overall result: Detected 11 Shift in Mean event(s) that were significant at the alpha
< 0.05 level
## # A tibble: 11 × 3
##    time_start time_end empirical_pvalue
##         <dbl>    <dbl>            <dbl>
##  1      75700    76200           0
##  2      75200    75700           0
##  3      74700    75200           0
##  4      74200    74700           0
##  5      76200    76700           0
##  6      76700    77200           0
##  7      77200    77700           0
##  8      78200    78700           0
##  9      73200    73700           0
## 10      73700    74200           0
## 11      77700    78200           0.0100
## Time uncertainty considered? TRUE Time ensemble supplied (n = 100)
## Paleo uncertainty considered? TRUE Values ensemble supplied (n = 1000)
## Error propagation ensemble members = 100
## Null hypothesis testing ensemble members = 100
## Members simulated using isospectral method
##
## Parameter choices:
## cpt.fun = changepoint::cpt.mean
## minimum.segment.length = 1
## method = AMOC
## penalty = MBIC
## ncpts.max = 1
```

Once again, we find some intervals with significant changes in the mean. Let's look at the plot to see how the timing of the abrupt shifts matches the ecological change.

```
MD03_plot <- plot(MD03_2707_ms)
```

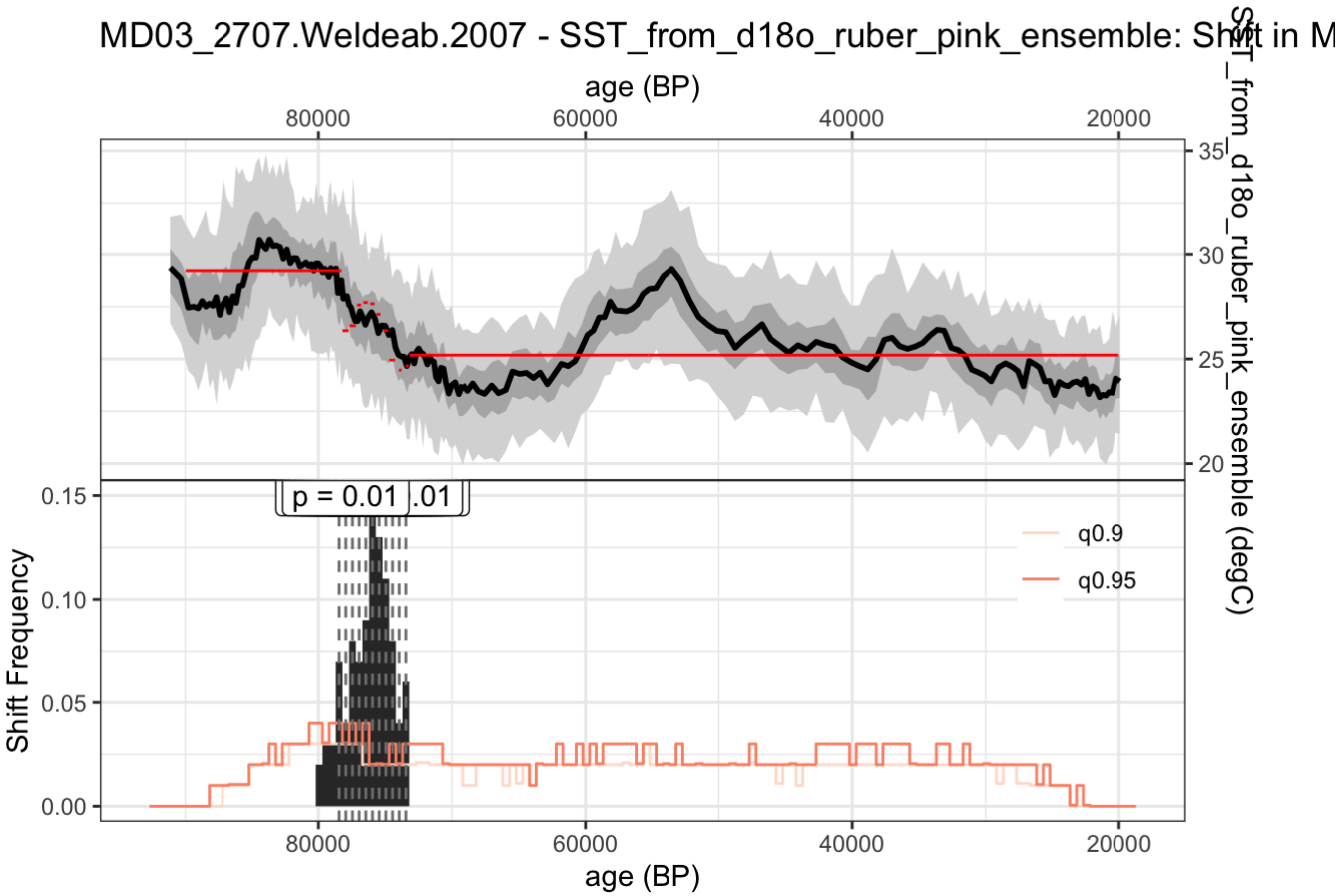

Figure 6. Results of the actR shift detection analysis for MDO3-2707 sea surface temperature estimates. Panels as in figure 4.

`actR::detectShift()` , with the parameters used here, consistently identifies a change in mean between 80,000 and 70,000 years ago, which suggests that the shift in Gulf of Guinea temperatures may have slightly proceeded or occurred simultaneously with the decrease in trees and shrubs inferred from Lake Bambili pollen.

To compare the results directly, we plot the likelihood of each of the shifts on the same figure.

```
ggplot() +
  geom_area(data = MD03_2707_ms$shiftDetection, aes(y = event_probability, x = time_mid,
fill = "Gulf of Guinea SSTs")) +
    geom_area(data = treeShrubMeanShift$shiftDetection, aes(y = event_probability, x = t
ime_mid, fill = "Lake Bambili Tree/Shrub percentage"), alpha = 0.8) +
  scale_fill_brewer("Dataset",palette = "Set1")+
  scale_x_reverse("Age (yr BP)") +
  ylab("Shift probability") +
  theme_bw() +
  ggtitle("Detected shifts in equatorial West Africa")
```

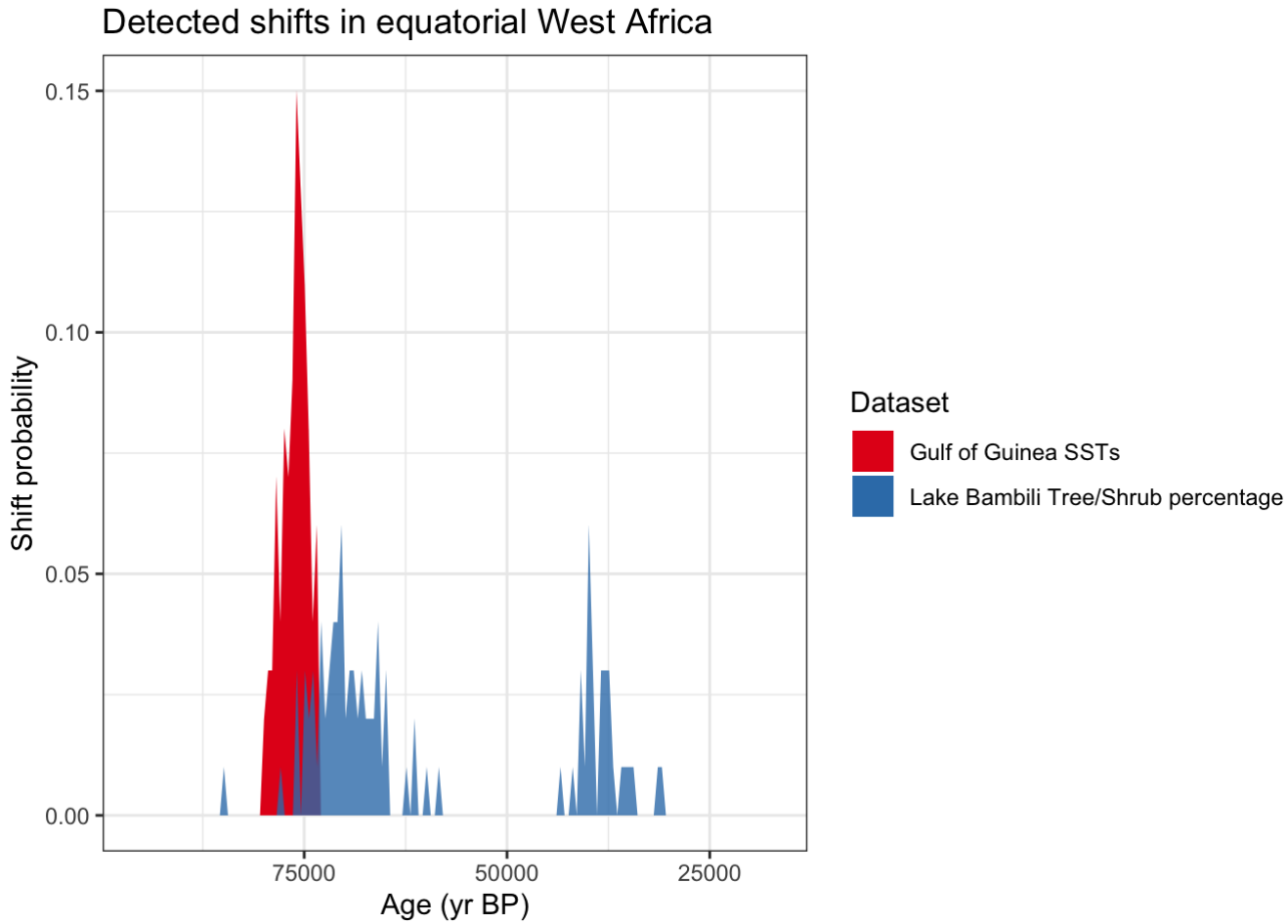

Figure 7. Detected shift frequency at Lake Bambili and in the Gulf Guinea. The fraction of ensemble members for which a shift was detected during a each 500 year window for each record, as shown in figures 4 and 6.

Clearly, there is more work to be done here to explore how changes in climate lead to community change in afromontane forests. This example simply highlights how these emerging tools can be used to efficiently and reproducibly work with paleogeoscientific data from multiple repositories.

Without these tools, preparing the data for analysis would be tedious, as the structure of the data can be heterogeneous. For example, without `neotoma2` and `neotoma2lipd()`, a user would access the Lake Bambili data by navigating the Neotoma Explorer website (https://apps.neotomadb.org/explorer/?datasetids=40944), finding Lake Bambili, and downloading csv files before manipulating for various analyses. The Gulf of Guinea SST data can be found at NOAA's World Data Service for Paleoclimatology (https://www.ncei.noaa.gov/pub/data/paleo/contributions_by_author/weldeab2007/weldeab2007.txt), and downloaded as a text file. The user would then edit that file as necessary before loading and manipulating for subsequent analysis.

References

Blaauw, M., and J. A. Christen. 2011. "Flexible Paleoclimate Age-Depth Models Using an Autoregressive Gamma Process." *Bayesian Analysis* 6 (3): 457–74.

Dominguez Vidana, Socorro, and Simon Goring. 2022. "Neotoma2 r Package." https://github.com/neotomadb/neotoma2 (https://github.com/neotomadb/neotoma2).

Goring, Simon. 2022. "Neotoma API 2.0." https://api.neotomadb.org/api-docs/ (https://api.neotomadb.org/api-docs/).

Heiser, Chris, and Nicholas McKay. 2022. “lipdR Package.” https://nickmckay.github.io/lipdR/ (https://nickmckay.github.io/lipdR/).

Kaufman, Darrell, Nicholas McKay, Cody Routson, Michael Erb, Basil Davis, Oliver Heiri, Samuel Jaccard, et al. 2020. “A Global Database of Holocene Paleotemperature Records.” *Scientific Data* 7 (1): 1–34.

Killick, Rebecca, and Idris A. Eckley. 2014. “changepoint: An R Package for Changepoint Analysis.” *Journal of Statistical Software* 58 (3): 1–19. http://www.jstatsoft.org/v58/i03/ (http://www.jstatsoft.org/v58/i03/).

Killick, Rebecca, Kaylea Haynes, and Idris A. Eckley. 2016. *changepoint: An R Package for Changepoint Analysis*. https://CRAN.R-project.org/package=changepoint (https://CRAN.R-project.org/package=changepoint).

Lézine, Anne-Marie, Kenji Izumi, Masa Kageyama, and Gaston Achoundong. 2019. “A 90,000-Year Record of Afromontane Forest Responses to Climate Change.” *Science* 363 (6423): 177–81.

McKay, Nicholas P, Julien Emile-Geay, and Deborah Khider. 2021. “geoChronR–an r Package to Model, Analyze, and Visualize Age-Uncertain Data.” *Geochronology* 3 (1): 149–69.

McKay, Nicholas, and Julien Emile-Geay. 2022. “Abrupt Change Toolkit in r.” http://linked.earth/actR/ (http://linked.earth/actR/).

Weldeab, Syee, David W Lea, Ralph R Schneider, and Nils Andersen. 2007. “155,000 Years of West African Monsoon and Ocean Thermal Evolution.” *Science* 316 (5829): 1303–7.

Williams, John W., Eric C. Grimm, Jessica L. Blois, Donald F. Charles, Edward B. Davis, Simon J. Goring, Russell W. Graham, et al. 2018. “The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource.” *Quaternary Research* 89 (1): 156–77. https://doi.org/10.1017/qua.2017.105 (https://doi.org/10.1017/qua.2017.105).