GeoChronR – an R framework to model and analyze age-uncertain paleogeoscientific timeseries.

Nicholas McKay[1], Julien Emile-Geay[2], and Deborah Khider[2]

[1]School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011
[2]University of Southern California, Los Angeles, CA

Correspondence: Nicholas McKay (Nicholas.McKay@nau.edu)

Abstract. Chronological uncertainty is a hallmark of the paleosciences. While many tools have been made available to researchers to produce age models suitable for various settings and assumptions, disparate tools and output formats often discourage integrative approaches. In addition, propagating age model uncertainties to subsequent analyses, and visualizing the results, has received comparatively little attention in the literature and available software. Here we describe GeoChronR, an open-source R package to facilitate these tasks. GeoChronR is built around emerging data standards for the paleosciences (Linked Paleo Data, or LiPD), and offers access to four popular age modeling techniques (Bacon, BChron, Oxcal, BAM). The output of these models is easily stored in LiPD, enabling paleo-aware synchronization, regression, correlation, principal component, and spectral analyses. Five real-world use cases illustrate how to use GeoChronR to facilitate these tasks, and to visualize the results in intuitive ways.

1 Introduction

1.1 Background

Quantifying chronological uncertainties, and how they might influence the understanding of past changes is fundamental to the paleogeosciences. Without robust error determination, it is impossible to properly assess the extent to which past changes occurred simultaneously across regions or the duration of abrupt events, both of which limit our capacity to apply paleoscientific understanding to modern and future processes. The need for better infrastructure to both characterize uncertainty, and to explicitly evaluate how age uncertainty impacts the the interpretation of records of past climate, ecology or landscapes, has been long recognized (Noren et al., 2013, (more)). In response to this need, the paleogeoscience community has made substantial advances toward improving geochronological accuracy by:

1. Improving analytical techniques that allow for more precise age determination on smaller and context-specific samples (**?**, **?**,**?**,**?**,**?**)

2. Refining our understanding of how past changes in the Earth system impact the age accuracy, for example: improvements to the radiocarbon calibration curve (**?**, **?**,intcal references) and advances in our understanding of spatial variability in cosmogenic production rates used in exposure dating (**?**, **?**)

3. Dramatic improvement in the level of sophistication and realism in age-depth models used to estimate the ages of sequences between dated samples (e.g. **?**, Parnell et al. (2008), **?**, **?**).

Over the past 20 years, these advances have been widely adopted in the paleogeosciences. However, despite the progress made in quantifying uncertainty in ages and in age models, few studies have formally evaluated how chronological uncertainty may have affected their results. For instance, whereas the algorithms presented by (**?**, Parnell et al. (2008), **?**, **?**) have been broadly used, the overwhelming majority of these studies calculate the single best-estimate model (often a median or mean), use this model to put measured paleoclimatic or paleoenvironmental data on a timescale, and then proceed to analyze the record with little to no reference to the age modeling exercise (**?**, **?**). Typically, any discussions of chronological uncertainties remain qualitative.

This paradigm is beginning to change. In recent years a handful of studies have taken advantage of approaches that generate ensembles of age models to evaluate how the results of their analyses and conclusions vary given differences between ensemble members (**?**, **?**,**?**,**?**,Marcott et al. (2013), **?**). By using each ensemble age model to create a time-uncertain ensemble records, and then carrying that ensemble through the analysis, the precise impact of age uncertainty can be formally evaluated. This approach, of course, does not address all aspects of uncertainty, but it does offer the broad potential to ascertain which results are robust to chronological uncertainty, and which are not.

Despite its potential to substantially improve uncertainty quantification for the paleogeosciences, this framework is not widely utilized. The majority of studies utilizing this approach have been regional [e.g., **?**,**?**,McKay et al. (2018), **?**} or global-scale (e.g., **?**, Marcott et al. (2013),**?**) syntheses. Occasionally, primary publications of new records incorporate time-unceratin analysis into their studies (**?**, more), but this remains rare. We suggest that there are several reasons for the lack of adoption of these techniques:

1. For synthesis studies, the necessary geochronological data are not publicly available for the vast majority of records. Even when they are available, the data are archived in diverse and unstructured data formats. Together, this makes what should be a simple process of aggregating and preparing data for analysis prohibitively time-consuming;

2. For studies of new and individual records, few tools for ensemble analysis are available, and those that are require a degree of comfort with coding languages and scientific programming that is rare among paleogeoscientists;

3. There is a disconnect between age-model development and time-uncertain analysis. Published approaches have utilized either simplified age-modeling approaches (**?**), or specialized approaches not used elsewhere in the community (**?**, **?**,Marcott et al. (2013),**?**,**?**).

Extracting the relevant data from commonly-used age-modelling algorithms, creating time-uncertain ensembles, then reformatting those data for analysis in available tools typically requires the development of extensive custom codesets.

1.2 Motivation

GeoChronR is built to lower the barriers to broader adoption of these emerging methods. It provides an easily-accessibly, open-source and extenisble software package of industry-standard and cutting-edge tools that provides users a single environment to create, analyse, and visualize time-uncertain data. GeoChronR is designed around emerging standards in the paleogeosciences that connects users to growing libraries of standardized datasets formatted in the Linked PaleoData format (**?**) and (**?**).

1.3 Outline of manuscript

This manuscript describes the design, analytical underpinnings and most common use cases of GeoChronR.

Section @ref(age-modeling) describes the integration of age modelling algorithms with GeoChronR.

Section @ref(age-uncertain-analysis) details the methods implemented for age uncertain analysis, and section @ref(use-cases) provides five real-world examples of how GeoChronR can be used for scientific workflows.

2 Age Modeling with GeoChronR

Context: Trachsel & Telford

2.1 Bacon

2.2 BChron

2.3 Oxcal

2.4 Banded Age Model (BAM)

3 Age-uncertain data analysis in GeoChronR

Intro on analytical approach, using ensembles. . . .

3.1 Correlation

3.2 Regression

3.3 Principal Component Analysis

3.4 Spectral Analysis

4 Visualization with GeoChronR

ggplot philosophy. Customization.

4.0.1 Timeseries

4.0.2 Geospatial

WTF does geospatial mean anyway? NM: It means mapping.

4.0.3 Spectral Analysis

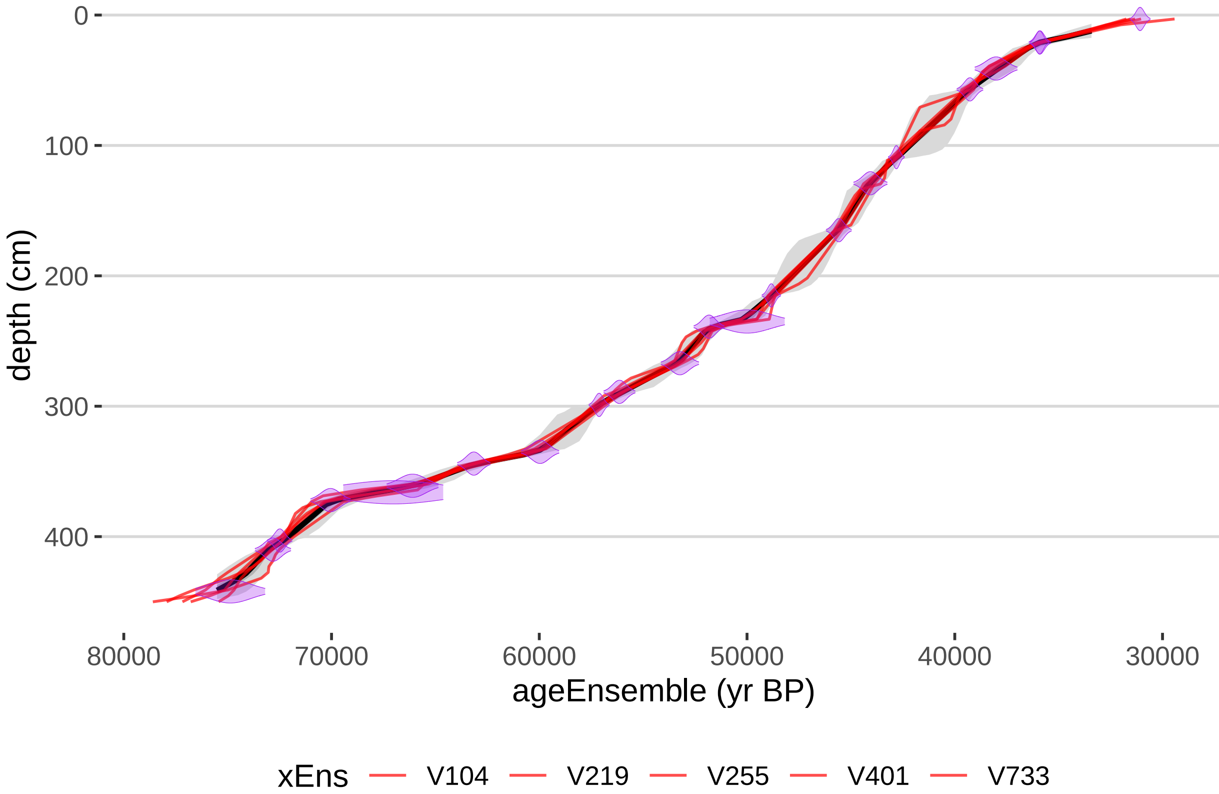
5 Use cases

We now illustrate the use of these tools on four use cases. The first example shows how to create age ensembles on different archives, and how to visualize the timing of abrupt events with appropriate uncertainty quantification. The second example shows how to quantify similarities in age-uncertain records. The third introduces the topic of age-uncertain calibrations, the fourth quantifies multivariate relationships using principal components analyses, and the fifth deal with spectral analysis.

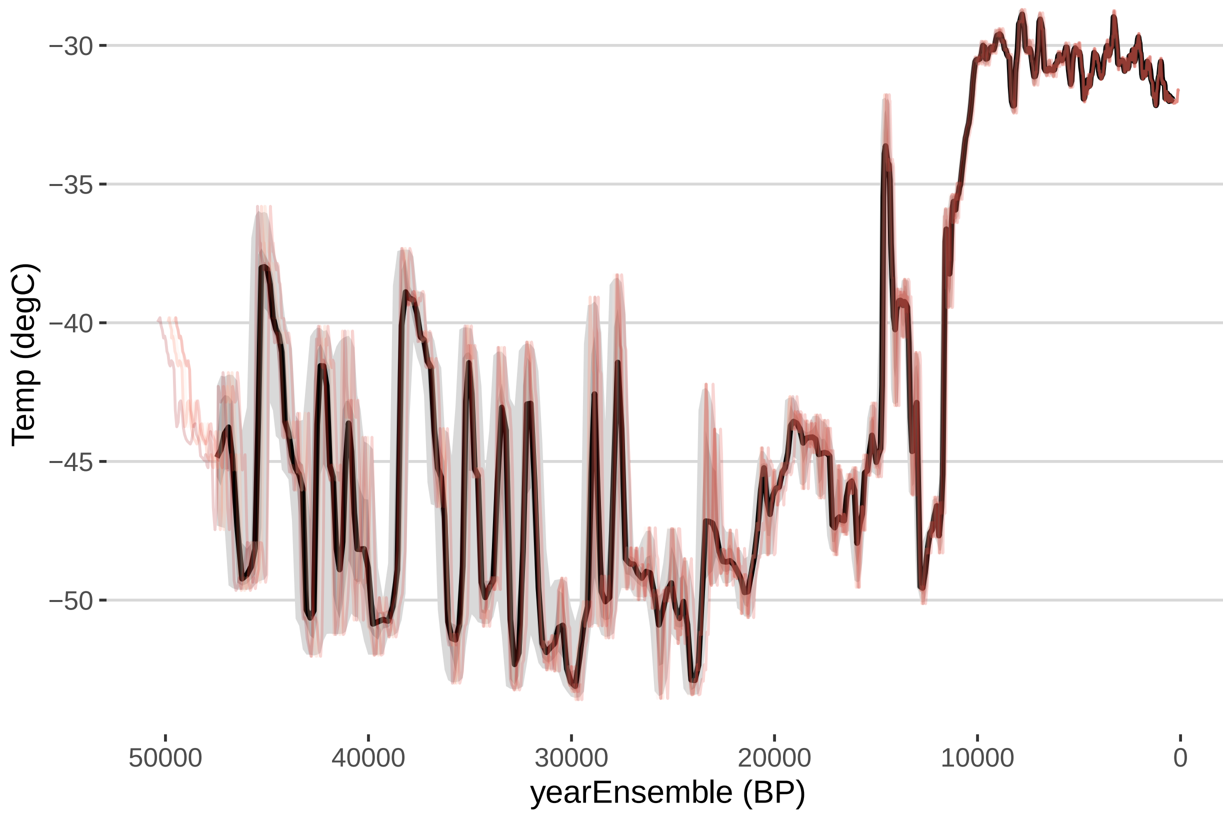
5.1 Creating an age ensemble

A common first task when using geoChronR is to create an age ensemble, either because the user is developing a new record, or because the age ensemble data for the record they are interested is unavailable. As described in section X.Y workflows for four published age quantification programs are integrated into geoChronR. Bacon (Blaauw and Christen, 2011), BChron (Parnell et al., 2008), and OxCal (Ramsey, 2008) are Bayesian age-deposition models that estimate posteriors on age-depth relationships with different assumptions and methodologies. BAM (Comboul et al., 2014) was designed to probabilistically simulate counting uncertainty in banded archives, such as corals, ice cores, or varved sediments, but can also be used to simulate age uncertainty for any record, and is useful when the data or metadata required to calculate an age-depth model are unavailable. All four methods are mostly simply used in geoChronR with a LiPD file that contains the chronological measurements, and the functions `runBacon(L)`, `runBchron(L)`, `runOxcal(L)` and `runBam(L)`. These functions take LiPD objects as inputs, and return updated LiPD objects that include age-ensemble data generated by the respective software packages. Typically, additional parameters are needed for to optimally run the algorithms. When these parameters are not included, geoChronR will run in interactive mode, asking the user which variables and parameters they would like to model. These parameter choices are printed to the screen during while the program runs, or are available later with the function `getLastVarString()`. By specifying these parameters, age model creation can be scripted and will run in non-interactive mode. In this use case, we'll use geoChronR and BChron (Parnell et al., 2008) to calculate an age ensemble for the Hulu Cave δ^{18}O speleothem record (**?**), and BAM (Comboul et al., 2014) to simulate age uncertainties for the GISP2 ice core δ^{18}O dataset (**?**). The `plotChronEns(hulu)` function will plot an age-depth model and uncertainties derived from the age ensemble.

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



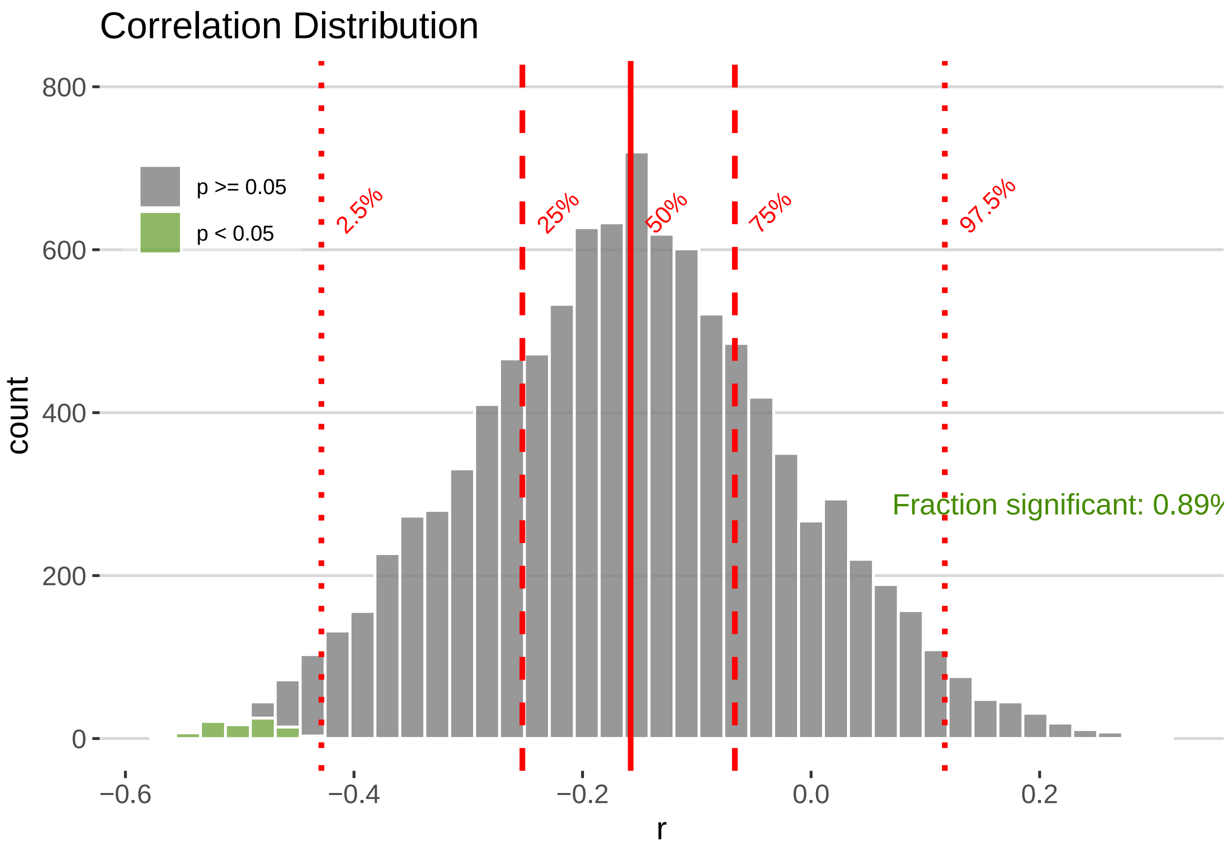
After an age ensemble has been added to a LiPD object, the user can visualize the ensemble timeseries using `plotTimeseriesEnsRibbons()` and `plotTimeseriesEnsLines()`. GISP2 δ^{18}O is plotted with age uncertainty, using both functions, in figure x.

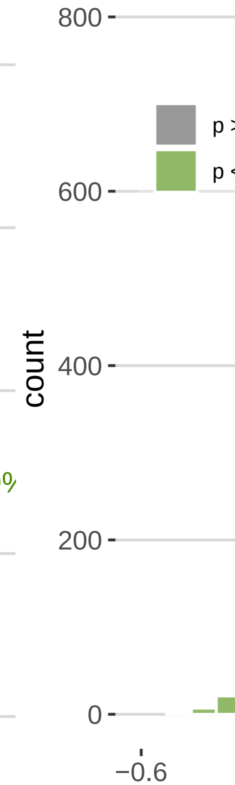


5.2 Correlation

Now that the user has generated age ensembles for the two datasets, they're interested to see if a correlation between the two datasets is robust to the age uncertainty modeled here. On multi-millennial timescales, the two datasets have similarities, and previous work has suggested that could events during the Last Glacial period, which are observed in the GISP2 record, can impact the Asian Monsoon and be observed is speleothem records such as the Hulu Cave dataset. (NM: add references here and flesh out background) The `corEns()` function in geoChronR will calculate ensemble correlations across age-uncertain datasets, such as these. `corEns()` will also sample across ensembles in the paleoData as well, if present. Here we calculate correlations during the period of overlap in 500 yr steps, determining significance for each pair of ensemble members while accounting for autocorrelation.

The results show consistently negative correlations, although 13.45% of the ensemble members are positive. However, only 0.89% are significant after accounting for serial autocorrelation.





In this use case, we demonstrate how a user could calculate and visualize and age-uncertain correlation between the Hulu speleothem δ^{18}O record and the GISP2 ice core δ^{18}O record. (Introduction about why a user might want to do an age uncertain correlation between Hulu and GISP2).

5.3 Age-uncertain Calibration

A natural extension of ensemble correlation is ensemble regression. Although there are use cases where regressing one age-uncertain variable onto another is called for, here we regress an age-uncertain paleoclimate proxy onto time-certain instrumental to develop a calibration-in-time. For this use case, we reproduce the results of Boldt, Brandon R. et al. (2015), where the authors calibrate a spectral reflectance measure of chlorophyll abundance to temperature in Northern Alaska.

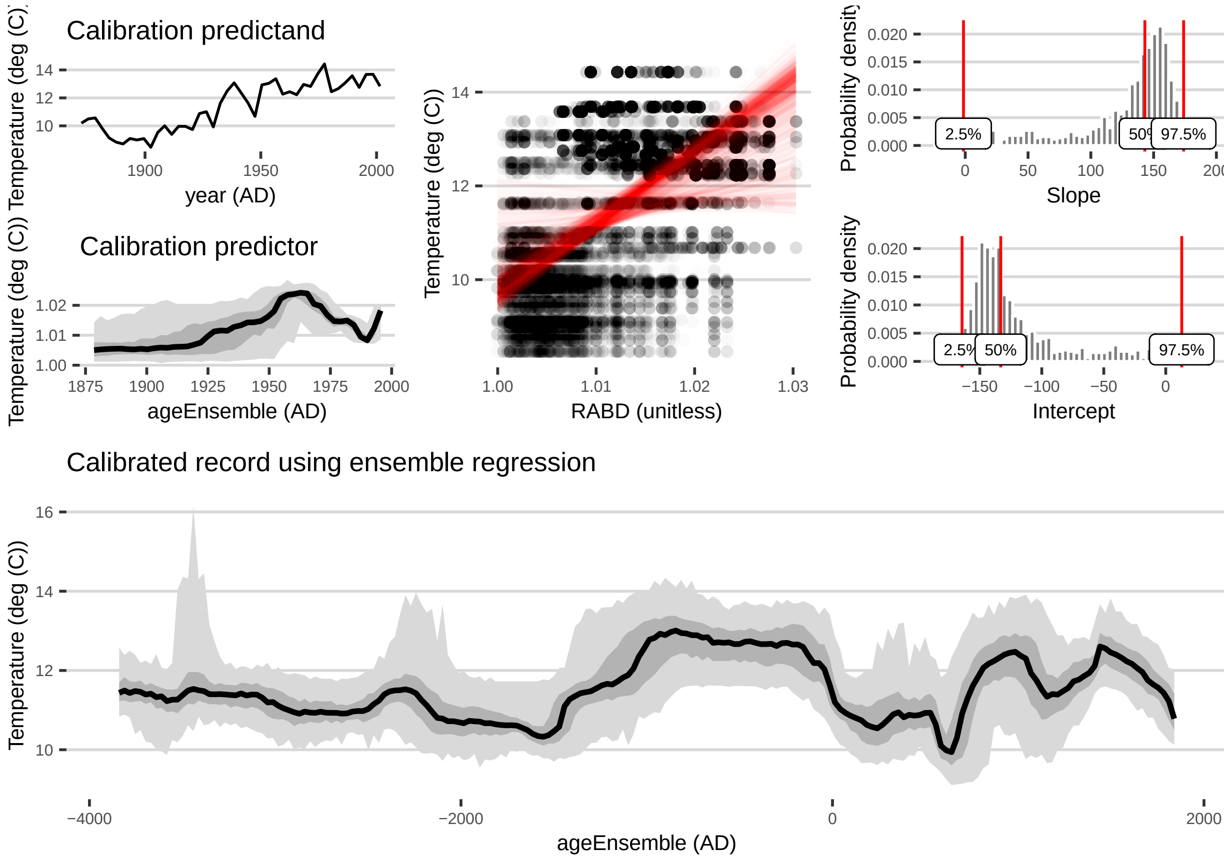
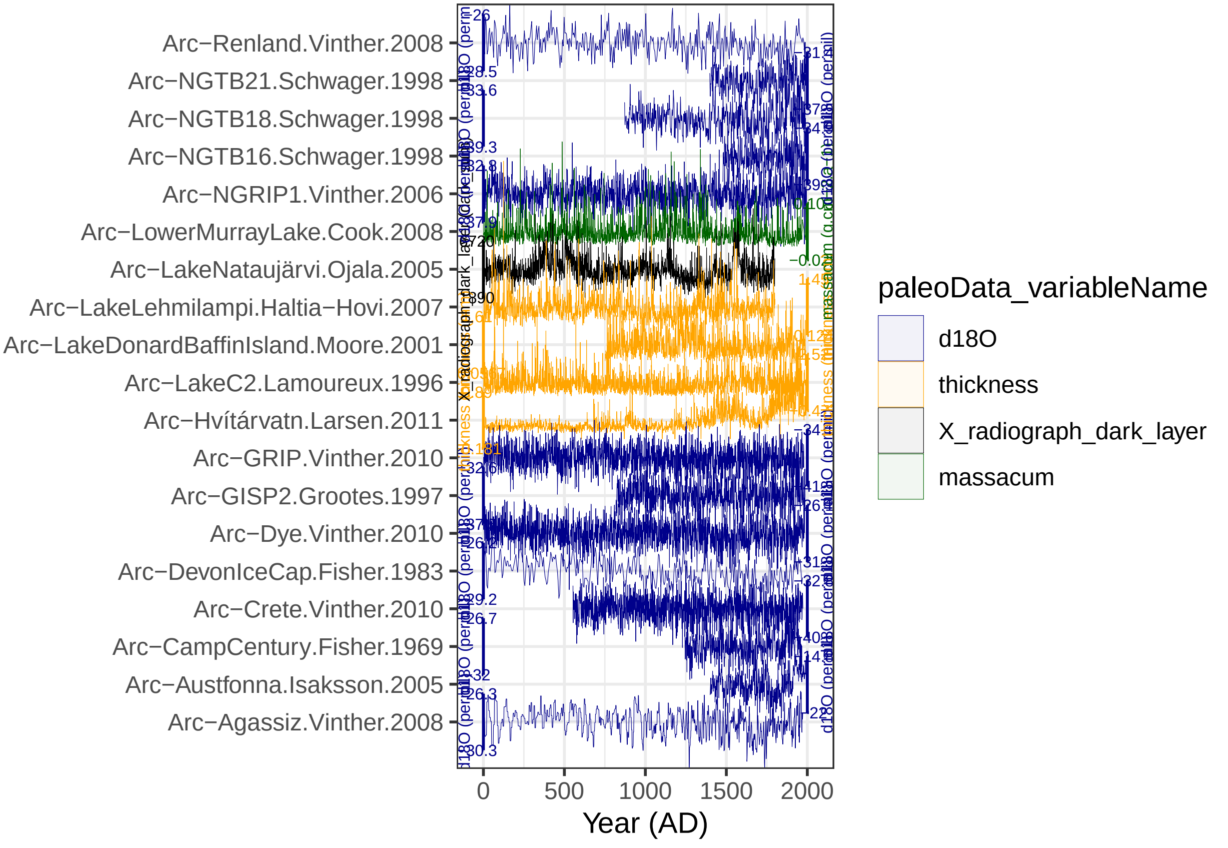
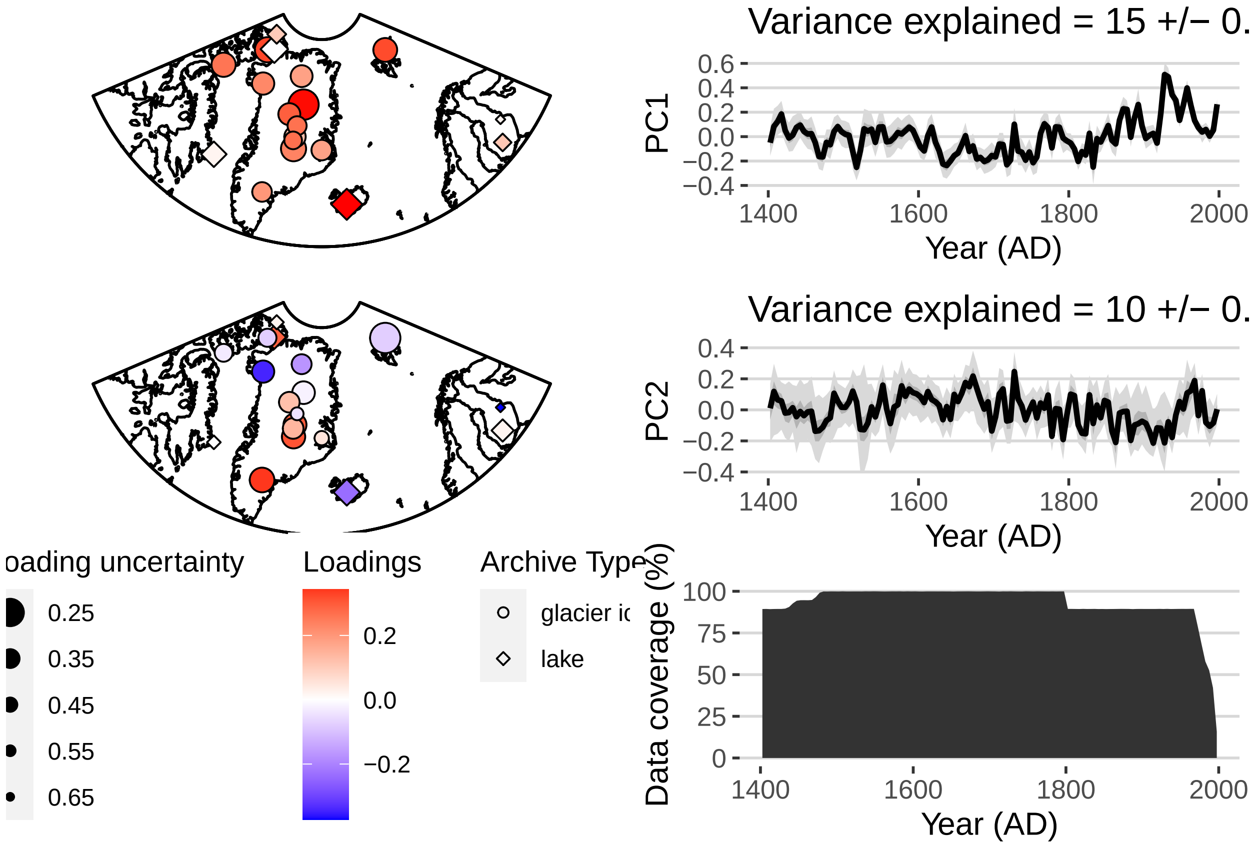


Figure captions.

5.4 Principle Component Analysis

Thus far, the use cases have highlighted age-uncertain analyses of one or two sites, however quantifying the effects of age uncertainty can be even more impactful over large spatial datasets. Here we showcase how to use geoChronR to perform age-uncertain principle components analysis (PCA), also known as Monte Carlo Empirical Orthogonal Function (MCEOF) analysis, pioneered by Anchukaitis and Tierney (2013). In this example, we examine the Arctic 2k database (**?**).


Arc−Renland.Vinther.2008
Arc−NGTB21.Schwager.1998
Arc−NGTB18.Schwager.1998
Arc−NGTB16.Schwager.1998
Arc−NGRIP1.Vinther.2006
Arc−LowerMurrayLake.Cook.2008
Arc−LakeNataujärvi.Ojala.2005
Arc−LakeLehmilampi.Haltia−Hovi.2007
Arc−LakeDonardBaffinIsland.Moore.2001
Arc−LakeC2.Lamoureux.1996
Arc−Hvítárvatn.Larsen.2011
Arc−GRIP.Vinther.2010
Arc−GISP2.Grootes.1997
Arc−Dye.Vinther.2010
Arc−DevonIceCap.Fisher.1983
Arc−Crete.Vinther.2010
Arc−CampCentury.Fisher.1969
Arc−Austfonna.Isaksson.2005
Arc−Agassiz.Vinther.2008
0
500
1000
1500
2000
Year (AD)
paleoData_variableName
d18O
thickness
X_radiograph_dark_layer
massacum



5.5 Spectral Analysis

6 Discussion & Conclusion

– Strengths, weaknesses and shortcomings of our approach

– Next steps; where does age-uncertain work go from here?

– GeoChronR plans, longevity, etc

7 Everything below are useful examples of how to use RMarkdown

Subsection text here.

7.0.1 Subsubsection Heading Here

Subsubsection text here.

8 Content section with citations

See the R Markdown docs for bibliographies and citations.

Copernicus supports biblatex. I put the .bib entries from the Paleocube proposal into `geochronr.bib`. Citations work like this:

Read (Evans et al., 2013), and (see Dee et al., 2015).

9 Content section with R code chunks

You should always use `echo = FALSE` on R Markdown code blocks as they add formatting and styling not desired by Copernicus. The hidden workflow results in 42.

You can add verbatim code snippets without extra styles by using ``` ``` ``` without additional instructions.

```
sum <- 1 + 41
```

10 Content section with list

If you want to insert a list, you must

- leave
- empty lines
- between each list item

```
- leave

- empty lines

- between each list item
```

11 Examples from the official template

11.1 FIGURES

When figures and tables are placed at the end of the MS (article in one-column style), please add

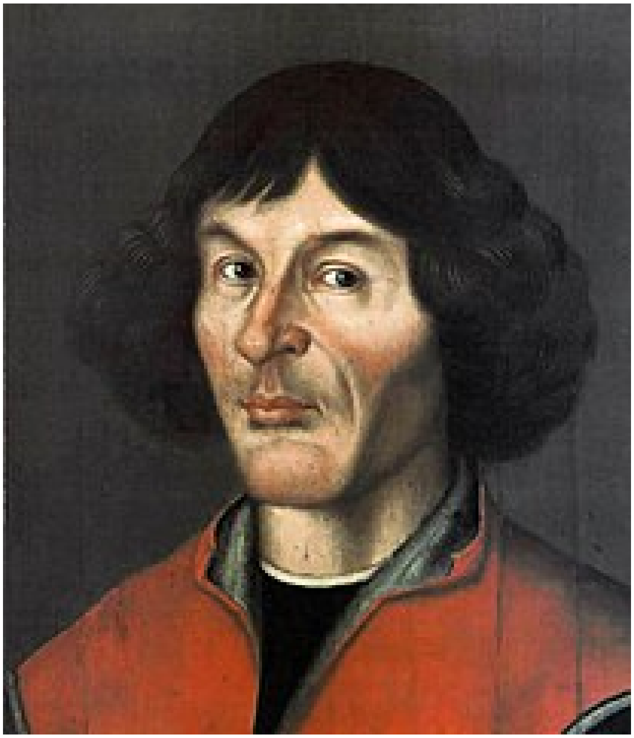


Figure 1. one column figure

between bibliography and first table and/or figure as well as between each table and/or figure.

11.1.1 ONE-COLUMN FIGURES

Include a 12cm width figure of Nikolaus Copernicus from Wikipedia with caption using R Markdown.

11.1.2 TWO-COLUMN FIGURES

You can also include a larger figure.

11.2 TABLES

You can ad LaTeXtable in an R Markdown document to meet the template requirements.

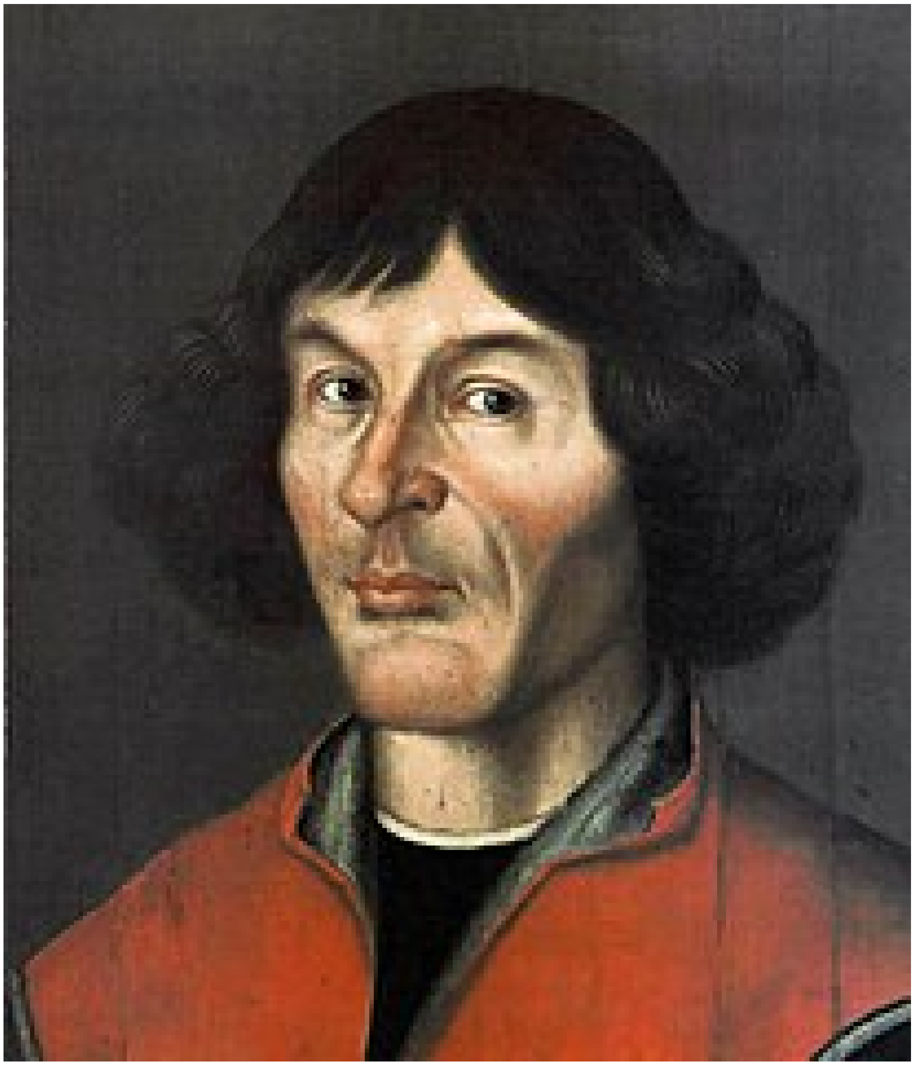


Figure 2. two column figure

Table 1. TEXT

a	b	c
1	2	3

Table Footnotes

Table 2. TEXT

a	b	c
1	2	3

Table footnotes

11.2.1 ONE-COLUMN TABLE

11.2.2 TWO-COLUMN TABLE

11.3 MATHEMATICAL EXPRESSIONS

All papers typeset by Copernicus Publications follow the math typesetting regulations given by the IUPAC Green Book (IUPAC: Quantities, Units and Symbols in Physical Chemistry, 2nd Edn., Blackwell Science, available at: http://old.iupac.org/publications/book 1993).

Physical quantities/variables are typeset in italic font (t for time, T for Temperature)

Indices which are not defined are typeset in italic font (x, y, z, a, b, c)

Items/objects which are defined are typeset in roman font (Car A, Car B)

Descriptions/specifications which are defined by itself are typeset in roman font (abs, rel, ref, tot, net, ice)

Abbreviations from 2 letters are typeset in roman font (RH, LAI)

Vectors are identified in bold italic font using $\boldsymbol{x}$

Matrices are identified in bold roman font

Multiplication signs are typeset using the LaTeX commands `\times` (for vector products, grids, and exponential notations) or `\cdot`

The character * should not be applied as mutliplication sign

11.4 EQUATIONS

11.4.1 Single-row equation

Unnumbered equations (i.e. using `$$` and getting inline preview in RStudio) are not supported by Copernicus.

$$1 \times 1 \cdot 1 = 42 \tag{1}$$

$$A = \pi r^2 \tag{2}$$

$$x = \frac{2b \pm \sqrt{b^2 - 4ac}}{2c}. \tag{3}$$

11.4.2 Multiline equation

$$3 + 5 = 8 \tag{4}$$
$$3 + 5 = 8 \tag{5}$$
$$3 + 5 = 8 \tag{6}$$

11.5 MATRICES

$$\begin{matrix} x & y & z \\ x & y & z \\ x & y & z \end{matrix}$$

11.6 ALGORITHM

If you want to use algorithms, you can either enable the required packages in the header (the default, see `algorithms: true`), or make sure yourself that the LaTeXpackages `algorithms` and `algorithmicx` are installed so that `algorithm.sty` respectively `algorithmic.sty` can be loaded by the Copernicus template. Copernicus staff will remove all undesirable packages from your LaTeX source code, so please stick to using the header option, which only adds the two acceptable packages.

11.7 CHEMICAL FORMULAS AND REACTIONS

For formulas embedded in the text, please use `\chem{}`, e.g. $\mathrm{A \rightarrow B}$.

The reaction environment creates labels including the letter R, i.e. (R1), (R2), etc.

Algorithm 1 Algorithm Caption

$i \leftarrow 10$
if $i \geq 5$ **then**
  $i \leftarrow i - 1$
else
  if $i \leq 3$ **then**
    $i \leftarrow i + 2$
  end if
end if

– `\rightarrow` should be used for normal (one-way) chemical reactions
– `\rightleftharpoons` should be used for equilibria
– `\leftrightarrow` should be used for resonance structures

$$A \rightarrow B \tag{R1}$$

$$Coper \rightleftharpoons nicus \tag{R2}$$

$$Publi \leftrightarrow cations \tag{R3}$$

11.8 PHYSICAL UNITS

Please use `\unit{}` (allows to save the math/`$` environment) and apply the exponential notation, for example $3.14\,\mathrm{km\,h^{-1}}$ (using LaTeX mode: `\( 3.14\,\unit{...} \)`) or $0.872\,\mathrm{m\,s^{-1}}$ (using only `\unit{0.872\,m\,s^{-1}}`).

12 Conclusions

The conclusion goes here. You can modify the section name with `\conclusions[modified heading if necessary]`.

Code and data availability. use this to add a statement when having data sets and software code available

Appendix A: Figures and tables in appendices

Regarding figures and tables in appendices, the following two options are possible depending on your general handling of figures and tables in the manuscript environment:

A1 Option 1

If you sorted all figures and tables into the sections of the text, please also sort the appendix figures and appendix tables into the respective appendix sections. They will be correctly named automatically.

A2 Option 2

If you put all figures after the reference list, please insert appendix tables and figures after the normal tables and figures.

To rename them correctly to A1, A2, etc., please add the following commands in front of them: `\appendixfigures` needs to be added in front of appendix figures `\appendixtables` needs to be added in front of appendix tables

Please add `\clearpage` between each table and/or figure. Further guidelines on figures and tables can be found below.

Competing interests. The authors declare no competing interests.

Disclaimer. disc

Acknowledgements. ack

References

Anchukaitis, K. J. and Tierney, J. E.: Identifying coherent spatiotemporal modes in time-uncertain proxy paleoclimate records, Climate dynamics, 41, 1291–1306, 2013.

Blaauw, M. and Christen, J.: Flexible paleoclimate age-depth models using an autoregressive gamma process, Bayesian Analysis, 6, 457–474, https://doi.org/doi:10.1214/11-BA618, 2011.

Boldt, Brandon R., Kaufman, Darrell S., McKay, Nicholas P., and Briner, Jason P.: Holocene summer temperature reconstruction from sedimentary chlorophyll content, with treatment of age uncertainties, Kurupa Lake, Arctic Alaska, The Holocene, https://doi.org/10.1177/0959683614565929, 2015.

Comboul, M., Emile-Geay, J., Evans, M., Mirnateghi, N., Cobb, K. M., and Thompson, D.: A probabilistic model of chronological errors in layer-counted climate proxies: applications to annually banded coral archives, Climate of the Past, 10, 825–841, https://doi.org/10.5194/cp-10-825-2014, 2014.

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D. M.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, Journal of Advances in Modeling Earth Systems, 7, 1220–1247, 2015.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Science Reviews, 76, 16–28, 2013.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of regional and global temperature for the past 11,300 years, Science, 339, 1198–1201, https://doi.org/10.1126/science.1228026, http://www.sciencemag.org/content/339/6124/1198.abstract, 2013.

McKay, Kaufman, D. S., Routson, C. C., Erb, M., and Zander, P. D.: The onset and rate of Holocene Neoglacial cooling in the Arctic., Geophysical Research Letters, https://doi.org/https://doi.org/10.1029/2018GL079773, 2018.

Noren, A., Brigham-Grette, J., Lehnert, K., Peters, S., Williams, J., Ito, E., Anderson, D., and Grimm, E.: Cyberinfrastructure for Paleogeoscience, workshop report, NSF EarthCube, 2013.

Parnell, A., Haslett, J., Allen, J., Buck, C., and Huntley, B.: A flexible approach to assessing synchroneity of past events using Bayesian reconstructions of sedimentation history, Quaternary Science Reviews, 27, 1872–1885, 2008.

Ramsey, C. B.: Deposition models for chronological records, Quaternary Science Reviews, 27, 42–60, 2008.