Supplementary Information

Long-term post-fire resilience of the Ericaceous Belt, Bale Mountains, Ethiopia

Graciela Gil-Romera, Carole Adolf, Blas M. Benito, Lucas Bittner, Maria U. Johansson, David A. Grady, Henry F. Lamb, Bruk Lemma, Mekbib Fekadu, Bruno Glaser, Betelhem Mekonnen, Miguel Sevilla-Callejo, Michael Zech, Wolfgang Zech, Georg Miehe

Contents

Note: The PDF version of this document only shows the most relevant code used in our analyses. The complete code can be found in the Rmd version, available at https://github.com/ggilromera/BaleFire . The analyses here presented were performed using the follwing packages: *png* [1], *grid*[2], *ggplot2*[3], *tidyr* [4], *viridis*[5], *nlme*[6], *cowplot*[7], *formatR*[8], and *knitr*[9]

1 Study site and core retrieval

Garba Guracha lake is 500 x 300 m (ca. 0,15 km^2) in size, with a maximum water depth of 6 m [10]. We recovered duplicate cores (BAL-GGU-1A, BAL-GGU-1B) using a Livingstone piston corer in February 2017, operated from a raft anchored at 5 m water depth. We reached a depth of 1680 cm (including water depth), similar to the depth previously published by Tiercelin et al. [1]. Both cores contain two sections between 12 and 13 m depth that could not be analysed as they are mainly coarse sand and gravel and could not be properly split when opening the cores or sampled. Further information on core retrieval can be found in Bittner et al.[11]

2 Pollen and charcoal sampling

We have analized samples between 1548 and 1472 cm and between 1178 and 60 cm, leaving out the unopen sections of the record. We analysed thus 1525 samples from which we are presenting in this study 1118, *i.e.* from 1178 to 60 cm. The top most 60 cm are still under analyses and the sediment below 1178 cm is not continuous while we want to focus our analyses on a continuous record of the fire and vegetation variables.

Charcoal particles (>150 μm) have been shown to be good proxies for local (*e.g.* Clark and Royall [12] or Peters and Higuera [13]) and sometimes even regional fire regimes [14,15]. Pollen source area (PSA) is both site and taxon-dependent as pollen productivity and dispersal patterns vary among plant species, and the size and nature of the lake and its catchment are approximately proportional to the pollen source area [16, 17]. *Erica* pollen grains are poorly dispersed, and their PSA has been proved to be essentially local [18,19].

Charcoal particles were sampled at contiguous intervals of core GGU1B (except in the coarse sand and gravel sections) and digested by soaking 1- 2 cm^3 sediment samples in 6% H_2O_2 for 48h, sieving at 150 μm and counting under a binocular microscope (×40). Charcoal counting was accomplished according to existing literature, counting opaque, angular particles [12]. We analysed and counted 275 samples of *Erica* fossil pollen grains using a modified version of the laboratory protocol of Moore (1981) and adding *Lycopodium* spores in a known number to estimate pollen accumulation rates (PAR) (Stockmarr, 1971). Both charcoal and pollen values were transformed to influx (accumulation rates measured as particles/cm^2 yr; CHAR and PAR respectively) in order 1) to account for the effect of different accumulation rates, 2) CHAR and PAR are better measurements of biomass burning and 3) to make them comparable when proceeding with numerical analyses.

3 Chronology: age model for BAL17-GGU-1B core

A full description of the methods used and the results obtained for the depth-age model of the GGU-IB core can be found in Bittner et al. [11]. The chronology is based on 24 radiocarbon-dated samples of bulk sediment, charcoal particles and n-alkanes from core BAL-GGU-1B and 23 samples from the surface core BAL-GGU-1A, dated by 210Pb-137Cs techniques. The age-depth model was built with the Bayesian approach implemented in the R package Bacon [20]. A summary of the samples dated and the model obtained can be found in ESM Table 1 and ESM Fig 1.

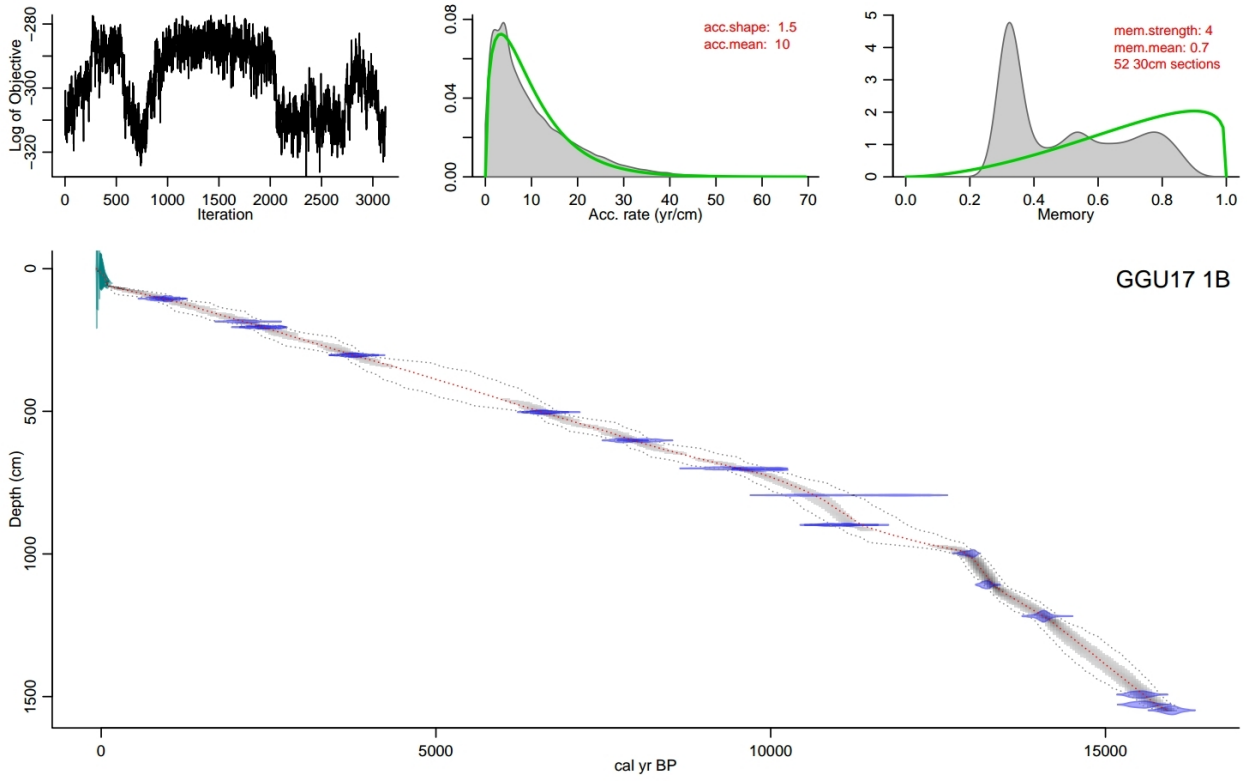

Figure 1: ESM Figure 1: Bayesian age-depth model for Garba Guracha lake (3950m asl) performed with the dates presented in ESM Table 1. More details are found in Bittner et al [2].

Lab. Name	Dating lab code	IPE-CSIC/ UH code	Dating method	Depth (cm)	Age (uncal. years BP)	Error	Age (cal. years BP)
ECRC -UCL	BAL0	BAL0	^{210}Pb-^{137}Cs	0	-67	1	
	BAL1.5	BAL1.5		1,5	-64	2	Not needed
	BAL4.5	BAL4.5		4,5	-47	2	
	BAL6.5	BAL6.5		6	-26	3	
	BAL7.5	BAL7.5		7	-13	5	
	BAL8.5	BAL8.5		8	-10	5	
	BAL9.5	BAL9.5		9	-8	5	
	BAL10.5	BAL10.5		10	-6	5	
	BAL12.5	BAL12.5		12	-1	5	
	BAL14.5	BAL14.5		14	4	5	
	BAL16.5	BAL16.5		16	8	6	
	BAL18.5	BAL18.5		18	13	6	
	BAL20.5	BAL20.5		20	17	6	
	BAL22.5	BAL22.5		22	21	7	
	BAL24.5	BAL24.5		24	26	7	
	BAL26.5	BAL26.5		26	31	8	
	BAL28.5	BAL28.5		28	34	9	
	BAL31.5	BAL31.5		31	40	10	
	BAL34.5	BAL34.5		34	46	11	
	BAL38.5	BAL38.5		38	56	13	
	BAL42.5	BAL42.5		42	68	14	
	BAL46.5	BAL46.5		46	78	15	
	BAL51.5	BAL51.5		51	94	18	
U.Bern	7931.1.1	GG1B1	^{14}C-BS	105	935	118	662-1081
	8273.1.1	GG1B1A	^{14}C-n-alkane	105	1076	79	892-1180
	8282.1.1	2L15.16	^{14}C-Charcoal	185	2124	129	1779-2366
	7930.1.1	GG1B2	^{14}C-BS	205	2323	111	2110-2722
	8272.1.1	GG1B2A	^{14}C-n-alkane	205	2399	101	2302-2743
	8271.1.1	GG1B3A	^{14}C-n-alkane	303	3476	89	3556-3979
	7929.1.1	GG1B3	^{14}C-BS	303	3517	111	3555-4091
	8270.1.1	GG1B5A	^{14}C-n-alkane	503	5789	109	6391-6804
	7928.1.1	GG1B5	^{14}C-BS	503	5794	135	6305-6903
	8269.1.1	GG1B6A	^{14}C-n-alkane	602	6967	123	7589-8003
	7927.1.1	GG1B6	^{14}C-BS	602	7320	144	7922-8404
	8268.1.1	GG1B7A	^{14}C-n-alkane	700	8267	137	8973-9535
	7926.1.1	GG1B7	^{14}C-BS	700	8753	156	9516-10201
	8279.1.1	7L35.36	^{14}C-Charcoal	705	8753	162	9496-10206
	7925.1.1	GG1B8	^{14}C-BS	794	10214	203	11267-12531
	8267.1.1	GG1B8A	^{14}C-n-alkane	794	9301	273	9740-11235
	8266.1.1	GG1B9A	^{14}C-n-alkane	898	9650	155	10545-11368
	7924.1.1	GG1B9	^{14}C-BS	898	9706	175	10563-11640
Direct AMS	D-AMS 029493	GG1B10	^{14}C-BS	998	11110	48	12828- 13082
	D-AMS 029494	GG1B11		1108	11377	50	13102- 13313
	D-AMS 029495	GG1B12		1218	12181	51	13906-14230
	D-AMS 027899	GG1B15		1493	12977	53	15291-15740
	D-AMS 029496	GG1B15b		1528	12997	57	15304-15772
	D-AMS 029497	GG1B15c		1548	13294	59	15772-16193

ESM Table 1: Dated samples in Garba Guracha using ^{210}Pb-^{137}Cs in core GGU-1A 1B and radiocarbon dating on different materials in core GGU-1B. Ages were calibrated using the INTCAL 13 2σ curve. More information is provided in Bittner et al. ECRC-UCL: Environmental Change Research Centre, Department of Geography, University College London. Ubern: Department of Chemistry and Biochemistry & Oeschger Centre for Climate Change Research, University of Bern; Direct-AMS: Radiocarbon Laboratory - WA 98011 USA. ^{14}C-BS: radiocarbon in bulk sediments; ^{14}C-n-alkane: radiocarbon in n-alkanes.

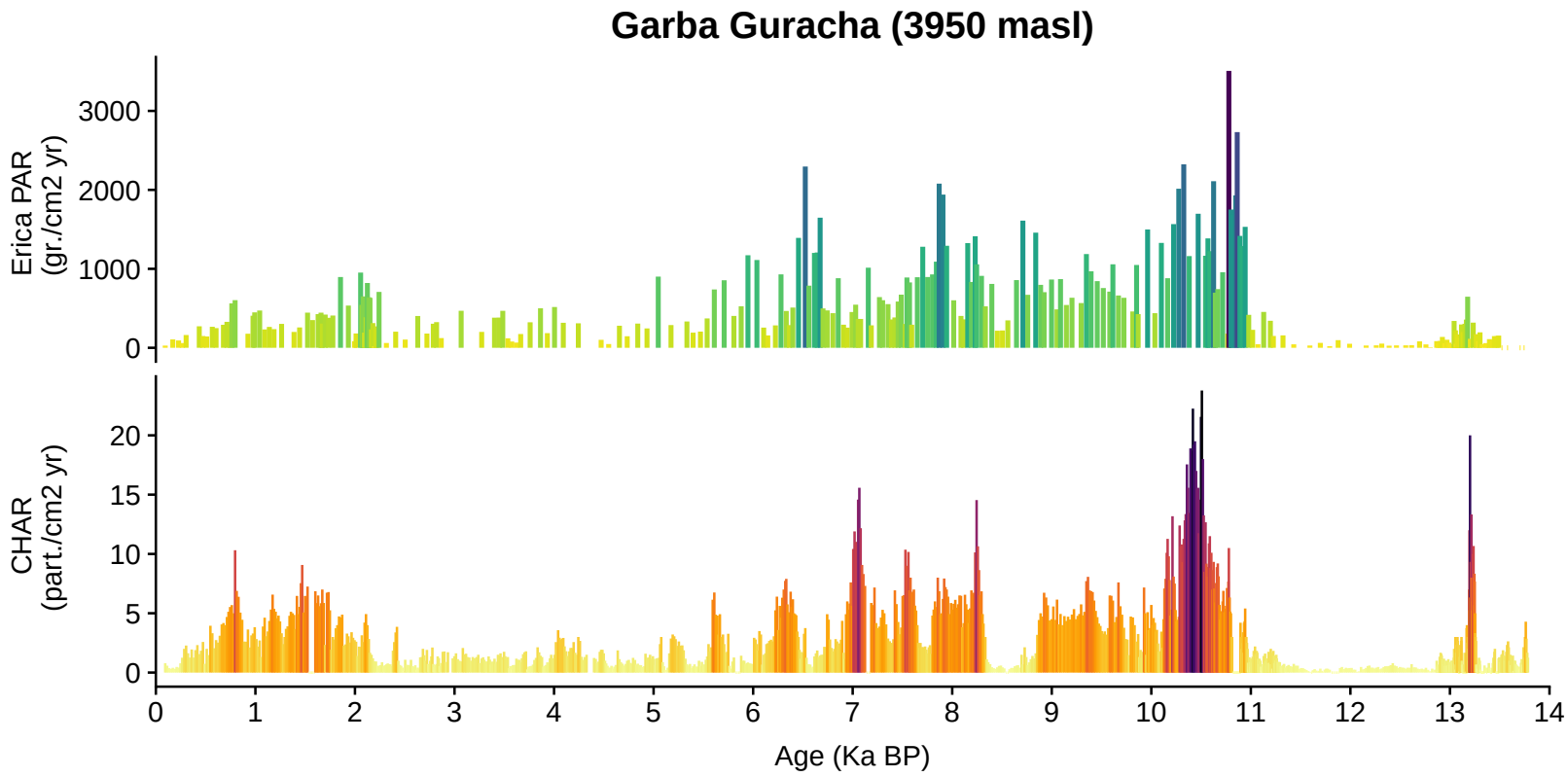

Figure 2: A) Erica PAR (grains/ cm2 year) B) CHAR (particles/ cm2 year)

4 Interpolation of charcoal into regular time intervals

To facilitate further analyses we interpolated the charcoal record into 10 years time-intervals. We modeled charcoal as a function of age with *loess*, by identifying the complexity value (parameter that controls de degree of smoothing) that maximized the correlation between observed and predicted charcoal values. We implemented the maximization algorithm within the function *interpolateDatasets*, which requires identifying the time span defining regular time intervals over which to interpolate the model result.

```
#requires time to find optimum span value
char.interpolated<-interpolateDatasets(
  datasets.list=list(char=char),
  age.column.name="age",
  interpolation.time.step=0.01 #ka
  )
```

The interpolation of charcoal into 10 years time-intervals only generated 248 new samples. To improve the quality of the interpolation, we replaced the interpolated values with the observed ones where possible, as shown below.

```
#matching decimal positions in ages
char.interpolated$age<-round(char.interpolated$age, 2)
char$age <- round(char$age, 2)

#for every age in char.interpolated
for(i in char.interpolated$age){

  #getting observed value for the given age
  observed.i <- char[char$age == i,"charcoal.acc.rate"]

  #if observed.i is not empty
  if(length(observed.i) > 0){
    #replacing interpolated by observed
    char.interpolated[
      char.interpolated$age == i,
      "charcoal.acc.rate"
      ] <- max(observed.i) #if two or more observed values with same age
  }
}
```

Figure 4 shows the comparison between observed and interpolated charcoal time-series.

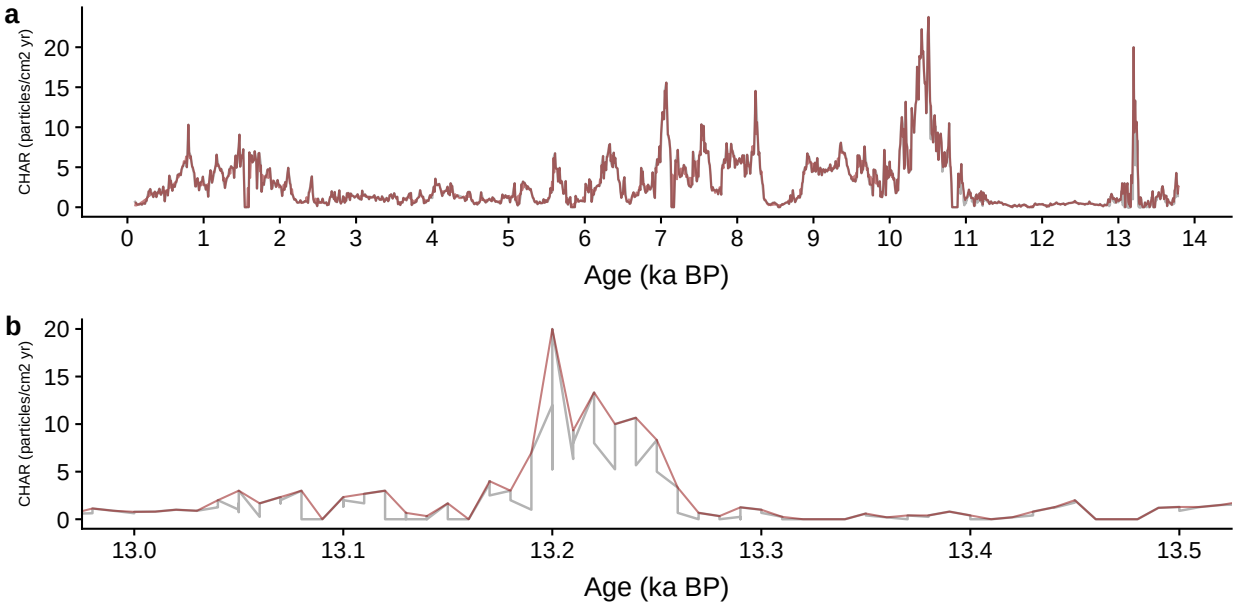

Figure 3: Comparison of observed (grey) and interpolated (red) charcoal time series. a) Complete time series, b) Zoom on a a particular time period where the interpolation partially decouples from real data

5 Generalised Least Squares

GLS provides robust estimates of regression parameters even when model residuals are heteroscedastic [21], a common problem of linear models fitted on time series data. We used the *gls* function of the **nlme** package [6] to fit three different sets of models, one assessing the relationship between synchronous values of charcoal and Erica PAR, and two others assessing time-delayed relationships between charcoal and Erica, and viceversa.

5.1 Synchronous model: concurrent effect of charcoal accumulation rate on Erica abundance

The code below shows the procedure used to fit the synchronous model (**Equation 1**), where Erica abundance was used as response variable, and charcoal accumulation rate (interpolated to regular time) as predictor. Intercept was left free, under the assumption that under zero fire, Erica abundances would likely be higher than zero.

Equation 1:

$$CHAR = \alpha + \beta PAR + \epsilon$$

Where:

- α represents the intercept.
- β is the coefficient estimate.
- ϵ is the error term.

```
#fitting GLS model
erica.char.gls <- gls(ericaceae.par ~ charcoal.acc.rate,
                      data=erica.char)

#pseudo R2 (gls doesn't provide R2)
erica.char.gls.R2 <- cor(erica.char$ericaceae.par,
    predict(erica.char.gls))^2
```

The model showed a pseudo R-squared value (*gls* does not compute R-squared) of 0.165. **Figure 4** shows the data and the fitted model.

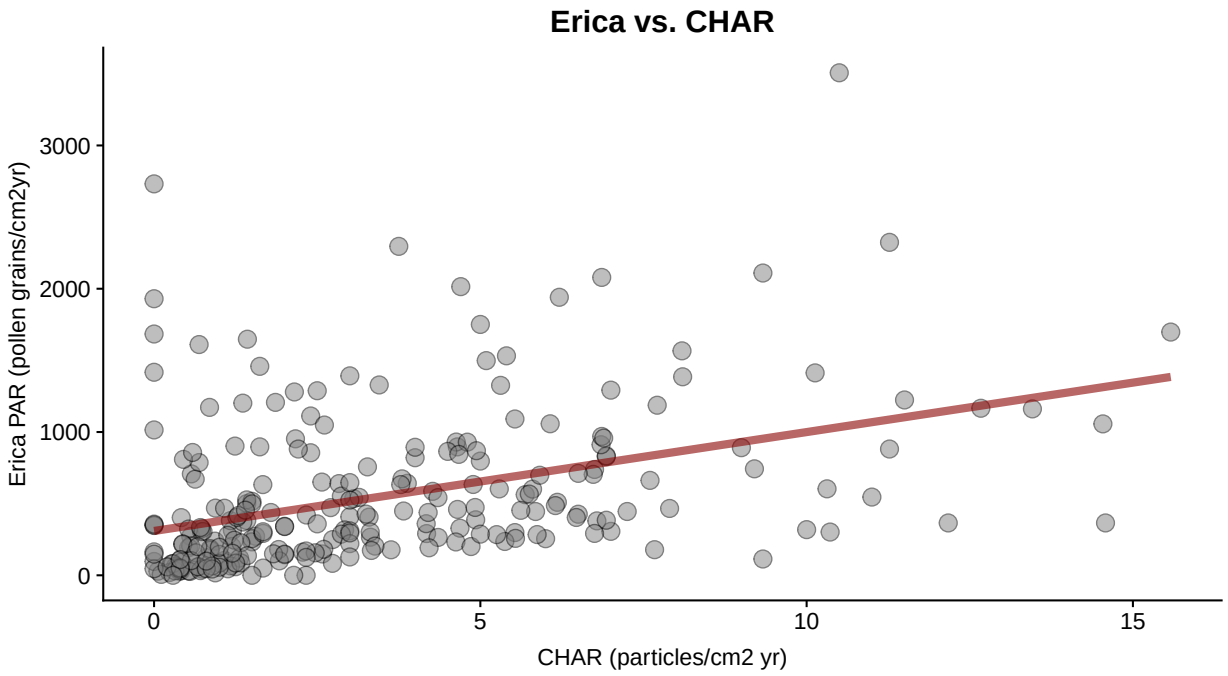

Figure 4: Erica abundances paired with synchronous samples of charcoal accumulation rate. Straight line shows the fit of the GLS model shown above..

5.1.1 Generating time-delayed (lagged) data

The analysis described in the next section requires the data to be expressed in time-lags, which involves aligning the samples of a given response variable with antecedent values of a given predictor or predictors. In our data, if we consider first Erica abundances (PAR) to be the response variable, and interpolated charcoal accumulation rates (CHAR) to be the predictor, for a lag of 10 years, the first PAR sample, with age 0.23 ka BP, has to be paired with the antecedent CHAR sample, with age 0.24 ka BP. The process is repeated for every sample and every lag, in our case up to 100 lags (1000 years) are considered.

For this study we generate two time-lagged datasets, one were Erica samples are paired with antecedent charcoal samples (named *lag.data.backward* in the code below), and one were charcoal samples are paired with antecedent Erica samples (named *lag.data.forward* in the code). We selected *backward* and *forward* as names because we consider *Erica* samples as reference. Therefore, the backward dataset is to assess the effect of "past" charcoal values on Erica, while the forward dataset is to assess the effect of Erica on "future" charcoal values.

The code below uses the custom functions **backwardsLags** and **forwardLags** to generate the datasets.

```
#100 lags
lags<-1:101

#forward dataset
lag.data.forward <- forwardLags(
  lags=lags,
  reference.data=erica,
  data.to.lag=char.interpolated
  )

#backward dataset
lag.data.backward <- backwardLags(
  lags=lags,
  reference.data=erica,
  data.to.lag=char.interpolated
  )
```

Figure 6 shows both datasets for lags 1 (10 years), 25 (250 years), 50 (500 years), 75 (750 years), and 100 (1000 years). It also shows an advance of the models that will be fitted in the following section.

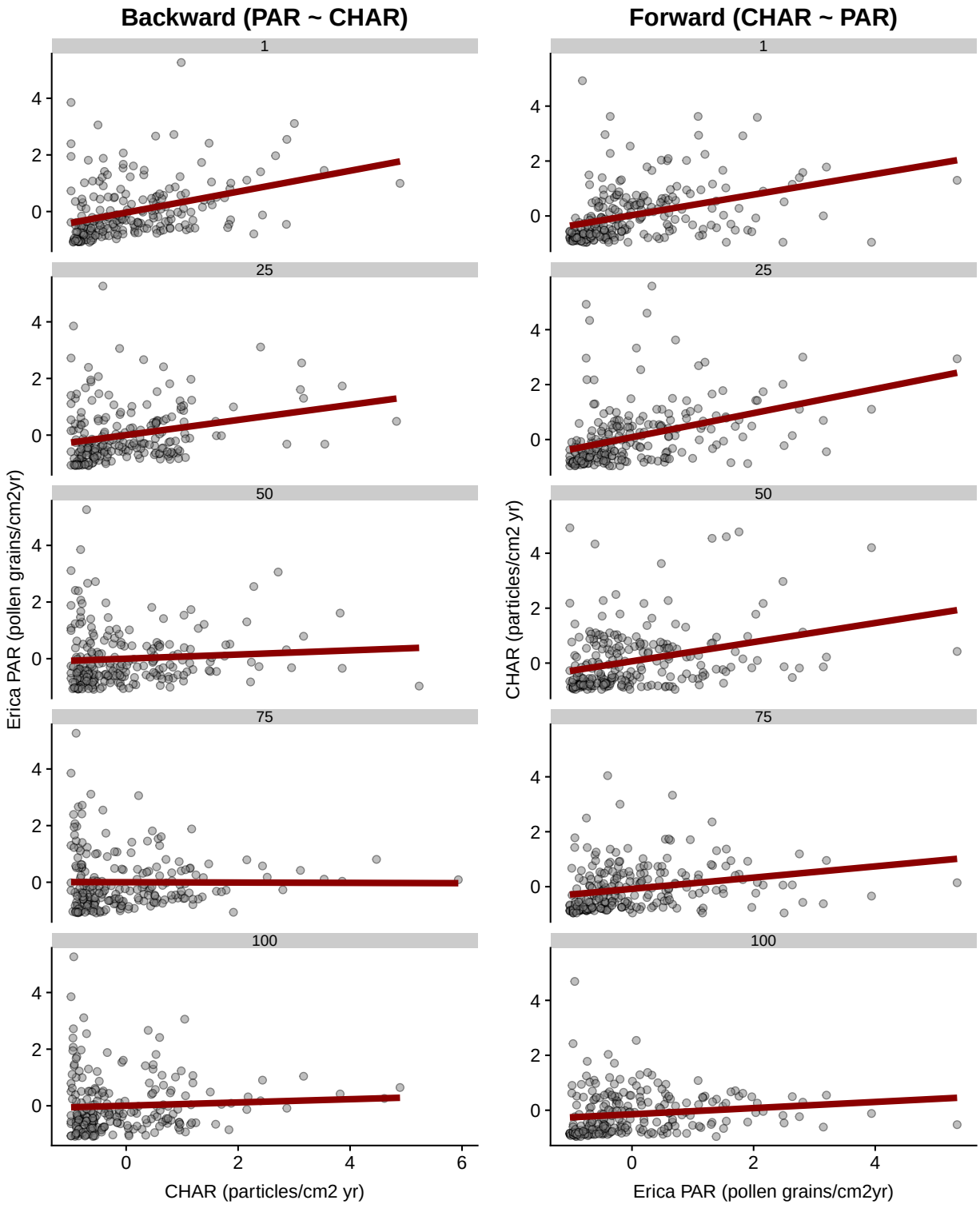

Figure 5: Lags 1, 25, 50, 75, , and 100 of the backward and forward datasets. Lag number can be found withing the grey strips. They have to be multiplied by 10 to convert lags into years. Lines represent linear models equivalent to those fitted in the following section.

5.2 Asynchronous models: time-delayed links between Erica abundance and charcoal accumulation rate

We aim to answer the questions: 1) Does Erica abundance have an effect on subsequent charcoal accumulation rates? 2) Are there time-delayed effects of charcoal accumulation rates on Erica abundances?;

To answer these questions we fitted two sets of *asynchronous* models described in Equations 2 and 3.

Equation 2:

$$CHAR_t = \alpha + \beta PAR_{t+lag} + \epsilon$$

Equation 3:

$$PAR_t = \alpha + \beta CHAR_{t+lag} + \epsilon$$

Where:

- t is the age of the given response sample.
- lag, with values between 10 and 1000 (in 10 years time-steps), represents the time-span in between the response samples and the antecedent samples of the given predictor.

Equations 2 and 3 were fitted once per lag on standardized data with generalised least squares (GLS) by using the gls function of the R package nlme [31]. Pseudo R-squared and standardized coefficient estimates with their respective confidence intervals were used to assess goodness of fit. 100 null models on permutated response variables were fitted for each equation to assess statistical significance.

We fitted the models through the custom function *modelLagData*, that requires a formula, and a lagged dataset. The function automatically fits one GLS model per lag, and stores pseudo R-squared values and standardized coefficient estimates.

```
#fitting a GLS model per lag on backward datasets
backward.results <- modelLagData(
  model.formula="ericaceae.par ~ charcoal.acc.rate",
  lagged.data=lag.data.backward
  )

#for the forward dataset
forward.results <- modelLagData(
  model.formula="charcoal.acc.rate ~ ericaceae.par",
  lagged.data=lag.data.forward
  )
```

To test the deviation of coefficient estimates and pseudo R-squared from random results we fitted each model 1000 times on permuted values of the response. The quantiles 0.05 and 0.95 of the resulting coefficient estimates and pseudo R-squared values were used as reference limits to differentiate random from non-random results. This operation was performed through the custom function *modelRandomLagData*, which takes a lagged dataset, a formula, and a number of iterations, and fits as many GLS models as iterations indicated, with the particularity that on each model the response is permutated. Aggregated pseudo R-squared values and standardized coefficient estimates of these models provide a robust null model to test the significance of our findings.

```
backward.results.random <- modelRandomLagData(
  lagged.data=lag.data.backward,
  model.formula="ericaceae.par ~ charcoal.acc.rate",
  iterations=1000
  )

forward.results.random <- modelRandomLagData(
  lagged.data=lag.data.forward,
  model.formula="charcoal.acc.rate ~ ericaceae.par",
  iterations=1000
  )
```

Coefficient estimates with their standard errors and pseudo R-squared values were extracted from each model and plotted against lagged age to facilitate the interpretation of the results. **Figure 7** shows the results of the fitted models.

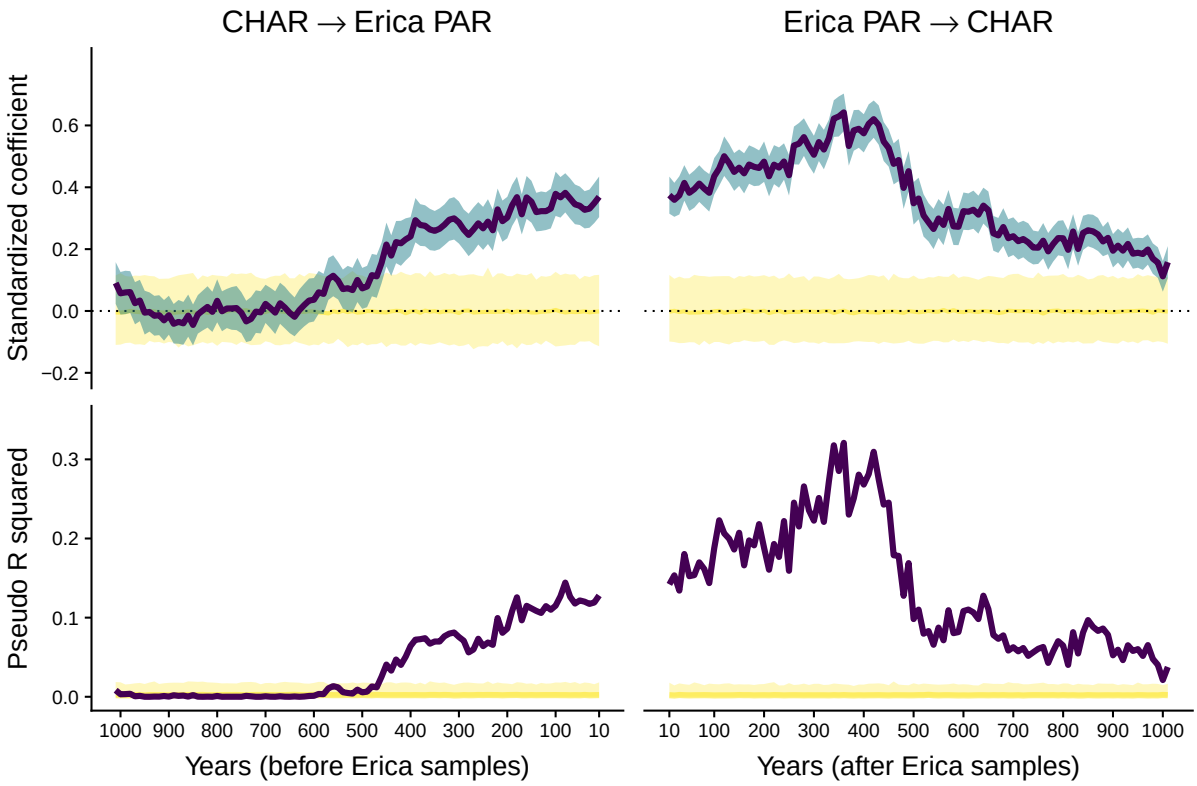

Figure 6: Results of the time-lagged models. Left panel represents models fitted with Equation 1, that is, the influence of antecedent values of CHAR on the pollen abundance of Erica (PAR) across time-lags up to 1000 years. The right panel represents the influence of antecedent values of Erica PAR on CHAR over the same time-lags. Yellow strips represent standardized coefficients and pseudo R-squared values for the null model. Data not intersecting yellow strips is interpreted as statistically significant.

6 Bibliography

[1] Urbanek, S. png: Read and write PNG images https://cran.r-project.org/web/packages/png/index.html . 2013.

[2] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. 2018.

[3] Wickham, H. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. http://ggplot2.org. 2016.

[4] Wickham, H. and Lionel, H. Tidyr: Easily Tidy Data with 'Spread()' and 'Gather()' Functions. https://CRAN.R-project.org/package=tidyr.2018.

[5] Garnier, Simon. Viridis: Default Color Maps from 'Matplotlib'. https://CRAN.R-project.org/package=viridis. 2018.

[6] Pinheiro J, et al. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137, https://CRAN.R-project.org/package=nlme. 2018.

[7] Wilke, C. Cowplot: Streamlined Plot Theme and Plot Annotations for 'Ggplot2'. https://CRAN.R-project.org/package=cowplot.2019.

[8] Xie, Y. FormatR: Format R Code Automatically. https://CRAN.R-project.org/package=formatR.2017.

[9] Xie, Y.Knitr: A General-Purpose Package for Dynamic Report Generation in R. https://yihui.name/knitr/. 2018.

[10] Tiercelin, J.J. et al., "High-resolution sedimentary record of the last deglaciation from a high-altitude lake in Ethiopia," Quaternary Science Reviews, vol. 27, no. 5, pp. 449–467, Mar. 2008.

[11] Bittner,L. et al. Revisiting Lake Garba Guracha in the afro-alpine Bale Mountains of Ethiopia - rationale, chronology and geochemistry. In prep.

[12] Clark, J.S. and Royall,P.D. Local and regional sediment charcoal evidence for fire regimes in presettlement north-eastern North America. Journal of Ecology, vol. 84, pp. 365–382, 1996.

[13] Higuera, P.E., et al. Understanding the origin and analysis of sediment-charcoal records with a simulation model. Quaternary Science Reviews, vol. 26, no. 13–14, pp. 1790–1809, 2007.

[14] Adolf, C. et al., The sedimentary and remote-sensing reflection of biomass burning in Europe. Global Ecol Biogeogr, vol. 27, no. 2, pp. 199–212, 2018.

[15] Oris, F. et al., 'Charcoal dispersion and deposition in boreal lakes from 3 years of monitoring:

Differences between local and regional fires', Geophysical Research Letters, vol. 41, no.2, p. 6743-6752. doi: 10.1002/2014GL060984, Aug. 2014.

[16] Davis, M.B. On the theory of pollen analysis. American Journal of Science, vol. 261, no. 10, pp 897–912. 1963.

[17] Davis, M.B. Redepostion of pollen grains in lake sediments. Limnology and Oceanography, vol. 18, no. 1, pp. 44–52, Jan. 1973.

[19] Schüler, et al., "Relationship between vegetation and modern pollen-rain along an elevational gradient on Kilimanjaro, Tanzania," The Holocene, vol. 24, no. 6, pp 702-713. 2014.

[19] Bonnefille, R. and Riollet,G. The Kashiru Pollen Sequence (Burundi) Palaeoclimatic Implications for the last 40,000 yr B.P. in Tropical Africa. Quaternary Research, vol. 30, pp. 19–35, 1988.

[20] Blaauw, M. and Christen,J.A. Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Analysis vol. 6, no. 3, pp. 457–474, 2011.

[21] Aitken,A.C. "IV.—On Least Squares and Linear Combination of Observations," Proceedings of the Royal Society of Edinburgh, vol. 55, pp. 42–48, 1936.