Reconstructing burnt area during the Holocene: an Iberian case study

Yicheng Shen[1,2,3], Luke Sweeney[1,2], Mengmeng Liu[3], Jose Antonio Lopez Saez[4], Sebastián Pérez-Díaz[5], Reyes Luelmo-Lautenschlaeger[4], Graciela Gil-Romero[6], Dana Hoefer[7], Gonzalo Jiménez-Moreno[8], Hieke Schneider[9], I. Colin Prentice[1,3], Sandy P. Harrison[1,2]

[1]Leverhulme Centre for Wildfires, Environment and Society, Imperial College London, South Kensington, London, SW7 2BW, UK
[2]Geography & Environmental Science, Reading University, Whiteknights, Reading, RG6 6AH, UK
[3]Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK
[4]Instituto de Historia, Centro de Ciencias Humanas y Sociales, Consejo Superior de Investigaciones Científicas, Madrid, Spain
[5]Department of Geography, Urban and Regional Planning, University of Cantabria, Santander, Spain
[6] Instituto Pirenaico de Ecología-CSIC, Avda. Montañana 1005, 50059, Zaragoza, Spain
[7] Senckenberg Research Station of Quaternary Palaeontology, Am Jakobskirchhof 4, 99423 Weimar, Germany
[8] Departamento de Estratigrafía y Paleontología, Facultad de Ciencias, Universidad de Granada, Avda. Fuente Nueva S/N, 18002 Granada, Spain
[9]Institut für Geographie, Friedrich-Schiller-Universität Jena, Löbdergraben 32, 07743 Jena, Germany

Correspondence to: Yicheng Shen (y.shen@reading.ac.uk)

This SI contains the following sections: (1) information about the preparation of the Iberia pollen data, (2) the analyses used to select the final GLM, (3) the justification of the use of the fourth root transformation of burnt area faction in the fxTWA-PLS analysis, (4) testing of the impact of using of micro and macro charcoal on the burnt area fraction reconstructions, and (5) comparisons of the fxTWA-PLS reconstructions with reconstructions based on alternative methods (WA-PLS, TWA-PLS).

This SI includes the following Figures and Tables:

Figure S1. Partial correlations test based on data sets for the Iberian Peninsula that include annual observations at 0.5° for the period between 2001 and 2016. The data are all transformed according to Table S2 prior to the test.

Figure S2. Partial residual plots of the final model (Table S3, last column). The plot includes (1) the expected value (blue line); (2) a confidence interval for the expected value (grey band); (3) partial residuals (dark grey dots).

Figure S3. Composite plots for macroscopic and macroscopic charcoal separately of 15 charcoal sites with both macroscopic and microscopic entities using the locfit() function with half-width=300, number of bootstrap samples=1000. The locally estimated scatterplot smoothing is shown in blue; The upper and lower 95th-percentile confidence intervals are shown in grey.

Figure S4. The fitted plots and residual plots of WA-PLS and TWA-PLS methods, with and without fx correction.

Figure S5. Composite curves of reconstructed burnt area using WAPLS and TWAPLS, with and without fx correction, using the locfit() function with half-width=300, number of bootstrap samples=1000. The locally estimated scatterplot smoothing is shown in blue; The upper and lower 95th-percentile confidence intervals are shown in grey.

Table S1. Pollen taxa in the Iberian dataset (205 taxa). Pollen taxa using in deriving the fire-vegetation relationship and fire reconstructions are shown in bold (139 taxa).

Table S2. Environmental variables and transformation methods

Table S3. The process of the model selection. Regression coefficients (t value) and pseudo-R^2 of each model are shown. The final model is shown in bold.

Table S4. Leave-out cross-validation fitness of WA-PLS and TWA-PLS methods, with and without fx correction, showing

S1. Preparation of the Iberian pollen data

Non-pollen palynomorphs (e.g. fungi, algae), introduced species, and fire-insensitive plants (e.g. obligate aquatics) were removed before analysis on the assumption that these were not diagnostic of changing fire regimes. Some pollen taxa are not identified consistently by palynologists or occur at very few sites, so some pollen types were amalgamated to higher taxonomic groups (genera for trees, families for herbaceous taxa) for consistency across the records (Table S1).

Table S1. Pollen taxa in the Iberian dataset (205 taxa). Pollen taxa using in deriving the fire-vegetation relationship and fire reconstructions are shown in bold (139 taxa).

Taxon:			
Abies	**Acer**	**Aconitum**	**Adonis**
Aesculus	Aizoaceae	**Alnus**	**Amaranthaceae**
Amaryllidaceae	Anacardiaceae	**Apiaceae**	Aquilegia
Araliaceae	**Arbutus**	Arctostaphylos	**Aristolochiaceae**
Artemisia	**Asparagaceae**	**Asphodelaceae**	**Asteraceae**
Asteraceae Liguliflorae	**Asteroideae**	**Astragalus**	**Berberidaceae**
Berberis	**Betula**	**Boraginaceae**	**Brassicaceae**
Buxus	Calicotome	**Calluna**	**Campanulaceae**
Caprifoliaceae	**Carduoideae**	**Carpinus betulus**	Carpinus orientalis Ostrya
Caryophyllaceae	**Castanea**	**Cedrus**	**Celastraceae**
Celtis	Ceratonia	Chamaerops	**Cichorioideae**
Cistaceae	**Cistus**	Clematis	**Colchicaceae**
Convolvulaceae	Coriaria	Cornus	**Corylus**
Crassulaceae	**Crataegus**	Cucurbitaceae	**Cupressaceae**
Cyperaceae	Cytinaceae	**Daphne**	Delphinium
Dennstaedtiaceae	Dryas	Elaeagnus	**Empetrum**
Ephedra	Ephedraceae	**Equisetum**	**Erica**
Ericaceae	Eriocaulaceae	**Euphorbiaceae**	**Fabaceae**
Fabaceae herbs	**Fagus**	**Frangula**	**Fraxinus**
Genisteae	**Gentianaceae**	**Geraniaceae**	Grossulariaceae
Halimium	Haloragaceae	**Hedera**	**Helianthemum**
Helleborus	Hippophae	**Huperzia**	Hymenophyllaceae
Hypericaceae	**Ilex**	**Iridaceae**	Jasminum
Juglans	**Juncaceae**	Juncaginaceae	Koenigia

Lamiaceae	Larix	**Lavandula**	Ledum
Ligustrum	**Liliaceae**	**Linaceae**	Linnaea
Linum	**Lonicera**	Loranthaceae	**Lycopodium**
Lysimachia	**Lythraceae**	**Malvaceae**	Maytenus
Melanthiaceae	Mercurialis	Montiaceae	**Moraceae**
Myrica	**Myrtaceae**	**Nartheciaceae**	Nerium
Nigella	**Olea**	**Oleaceae**	**Onagraceae**
Ononis	**Ophioglossaceae**	Orchidaceae	**Orobanchaceae**
Osmundaceae	Oxalidaceae	**Oxyria Rumex**	**Paeonia**
Papaveraceae	Parrotia	Periploca	**Phillyrea**
Picea	**Pinus**	**Pinus diploxylon**	Pinus haploxylon
Pistacia	**Plantaginaceae**	Platanus	**Plumbaginaceae**
Poaceae	**Polemoniaceae**	**Polygalaceae**	**Polygonaceae**
Polygonum	**Polypodiales**	**Populus**	Portulacaceae
Potentilla	**Primulaceae**	**Prunus**	**Pteridaceae**
Pterocarya	**Quercus deciduous**	**Quercus evergreen**	**Quercus intermediate**
Ranunculaceae	**Ranunculus**	**Resedaceae**	Retama
Rhamnaceae	**Rhamnus**	Rhododendron	Rhus
Ribes	**Rosaceae**	Rosmarinus	**Rubiaceae**
Rubus	Rutaceae	**Salix**	**Sambucus**
Sanguisorba group	Santalaceae	Sapotaceae	**Saxifragaceae**
Sciadopityaceae	**Scrophulariaceae**	Smilax	**Solanaceae**
Sorbus	**Tamarix**	Taxaceae	**Taxus**
Teucrium	**Thalictrum**	**Thymelaeaceae**	**Tilia**
Trollius	**Ulmus**	Ulmus Zelkova	**Urticaceae**
Vaccinium	**Valerianaceae**	Verbenaceae	**Viburnum**
Violaceae	Viscum	Vitex	**Ziziphus**
Zygophyllaceae			

S2. Selection of the final GLM

We initially examined 13 variables that have been shown in global analyses to influence burnt area. Table S2 provides information on the source of each of the data sets, their original resolution, and the transformation used prior to analysis.

Table S2. Environmental variables and transformation methods.

Environmental variables	Data source	Resolution	Transformation	Reference
Dry days per month	CRUNCEP V7	0.5° × 0.5°	Logarithmic	Viovy, 2018
Diurnal temperature range (K)	CRUNCEP V7	0.5° × 0.5°	Logarithmic	Viovy, 2018
Maximum temperature (K)	CRUNCEP V7	0.5° × 0.5°	Logarithmic	Viovy, 2018
Wind speed (m/s)	CRUNCEP V7	0.5° × 0.5°	Logarithmic	Viovy, 2018
Gross primary production (gC m^{-2} day^{-1})	FLUXCOM	0.5° × 0.5°	Logarithmic	Jung et al., 2020
Tree cover (%)	VCF	0.05° × 0.05°	Cell fraction	Hansen and Song, 2018
Non-tree cover (%)	VCF	0.05° × 0.05°	Cell fraction	Hansen and Song, 2018
Cropland (km^2)	HYDE 3.2	0.083° × 0.083°	Cell fraction	Klein Goldewijk et al., 2017
Total rainfed other crops (no rice) (km^2)	HYDE 3.3	0.083° × 0.083°	Cell fraction	Klein Goldewijk et al., 2017
Grazing land (km^2)	HYDE 3.2	0.083° × 0.083°	Cell fraction	Klein Goldewijk et al., 2017
Total population density (inhabitants km^{-2})	HYDE 3.2	0.083° × 0.083°	Square root	Klein Goldewijk et al., 2017
Urban population density (inhabitants km^{-2})	HYDE 3.2	0.083° × 0.083°	Square root	Klein Goldewijk et al., 2017
Rural population density (inhabitants km^{-2})	HYDE 3.2	0.083° × 0.083°	Square root	Klein Goldewijk et al., 2017

The process of final model selection: The partial correlation test (Fig. S2) shows that tree cover is highly correlated with both GPP (0.63) and non-tree cover (0.69), and rainfed cropland is highly correlated with total cropland (0.62). The three population variables are also strongly correlated with one another (> 0.80). There are moderate correlations between maximum temperature of the warmest month, maximum diurnal temperature range, and maximum dry days per month (> 0.40). We tested the impact of including or removing highly and moderately correlated variables before selecting the final model. Tree cover was not included in any GLM model because of its high correlation with both GPP and non-tree cover. The GLM model including cropland has a higher pseudo-R^2 than the model including total rainfed cropland (Table S3: first 2 columns), so only total cropland was retained. Comparison of the GLM models using total, urban and rural population density (Table S3: 2-4 columns) shows that only urban population density is statistically significant and the model with urban population density has the best fit (pseudo-R^2=0.20). All the variables in this model, except for maximum temperature, are statistically significant ($P < 0.1$). Given the lack of statistical significance of maximum temperature and the moderate correlations between maximum temperature and both diurnal temperature range and dry days per month (Fig. S2), we constructed three models leaving out one variable in turn. We obtained pseudo-R^2 values of 0.20, 0.17 and 0.20 respectively for these three models (Table S3: the last 3 columns). The model which does not include maximum temperature has the best fit (pseudo-R^2= 0.20). The final model was constructed using eight variables (Table S3: the last column).

Figure S1. Partial correlations test based on data sets for the Iberian Peninsula that include annual observations at 0.5° for the period between 2001 and 2016. The data are all transformed according to Table S2 prior to the test.

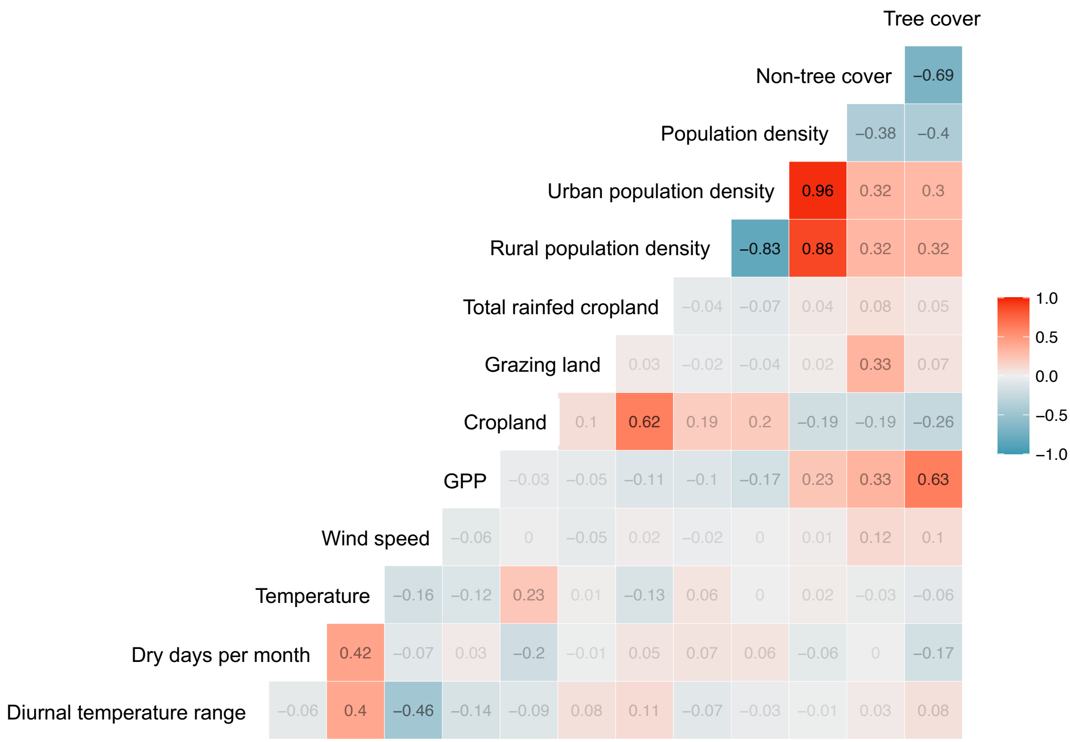


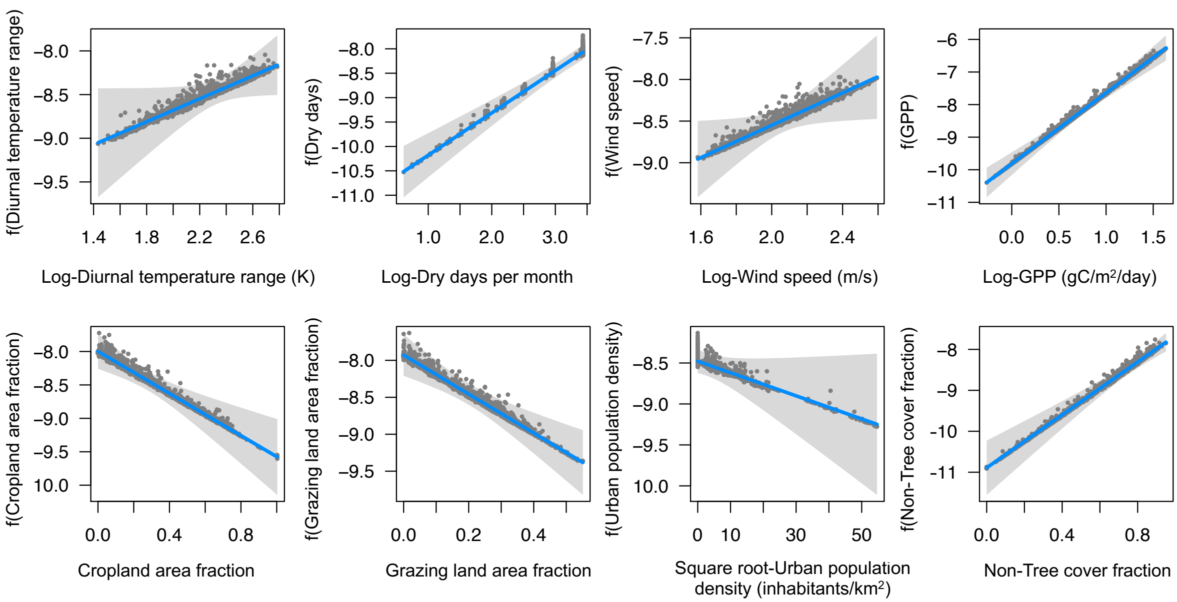
Table S3. The process of the model selection. Regression coefficients (t value) and pseudo-R^2 of each model are shown. The final model is shown in bold.

	1	2	3	4	5	6	**7**
	Include total rainfed cropland	Include total population density	Include rural population density	Include urban population density	Remove diurnal temperature range	Remove dry days per month	**Remove maximum temperature**
Diurnal temperature range	$1.95^{\cdot}$	$1.86^{\cdot}$	2.38^{*}	$1.77^{\cdot}$		1.19	$\mathbf{1.90}^{\cdot}$
Dry days per month	7.81^{***}	7.66^{***}	7.56^{***}	7.66^{***}	-1.01^{***}		$\mathbf{8.46}^{***}$
Maximum temperature	-0.50	0.01	-0.19	0.07	7.60	4.07^{***}	
Wind speed	2.00^{*}	2.04^{*}	2.05^{*}	2.07^{*}	0.69	$1.89^{\cdot}$	$\mathbf{2.11}^{*}$
GPP	9.54^{***}	9.76^{***}	9.49^{***}	9.85^{***}	1.43^{***}	9.41^{***}	$\mathbf{10.10}^{***}$
Non-tree cover	7.11^{***}	7.27^{***}	7.19^{***}	7.33^{***}	7.44^{***}	9.50^{***}	$\mathbf{7.34}^{***}$
Cropland		-3.95^{***}	-4.08^{***}	-3.95^{***}	9.76^{***}	-4.86^{***}	$\mathbf{-4.04}^{***}$
Total rainfed cropland	-3.62^{***}						
Grazing land	-4.64^{***}	-4.29^{***}	-4.19^{***}	-4.36^{***}	-4.17^{***}	-4.96^{***}	$\mathbf{-4.36}^{***}$
Total population density	-1.55	-1.07					
Rural population density			0.06				
Urban population density				$-1.68^{\cdot}$	-2.32^{*}	-1.38	$\mathbf{-1.69}^{\cdot}$
Pseudo-R^2	0.2008	0.2031	0.2013	0.2020	0.2012	0.1654	**0.2031**

Notes: $\cdot p < 0.1$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

Figure S2 shows the partial residuals obtained for the final model.

Figure S2. Partial residual plots of the final model (Table S3, last column). The plot includes (1) the expected value (blue line); (2) a confidence interval for the expected value (grey band); (3) partial residuals (dark grey dots)



S3. Justification for the use of the fourth root of the palaeo burnt area fraction used in the fxTWA-PLS analyses

Fire accelerates as it spreads. During the time between $t = 0$ (when it starts) and $t = t_1$ (when it stops), the total distance covered by the fire front is given by:

$$Total\ distance = \int_0^{t_1} ROS\ (t) \times dt$$

where ROS (t) is the rate of spread as a function of t. If $ROS\ (t) = a \times t$, then this integral is given by:

$$Total\ distance = \frac{a}{2} \times t_1^2$$

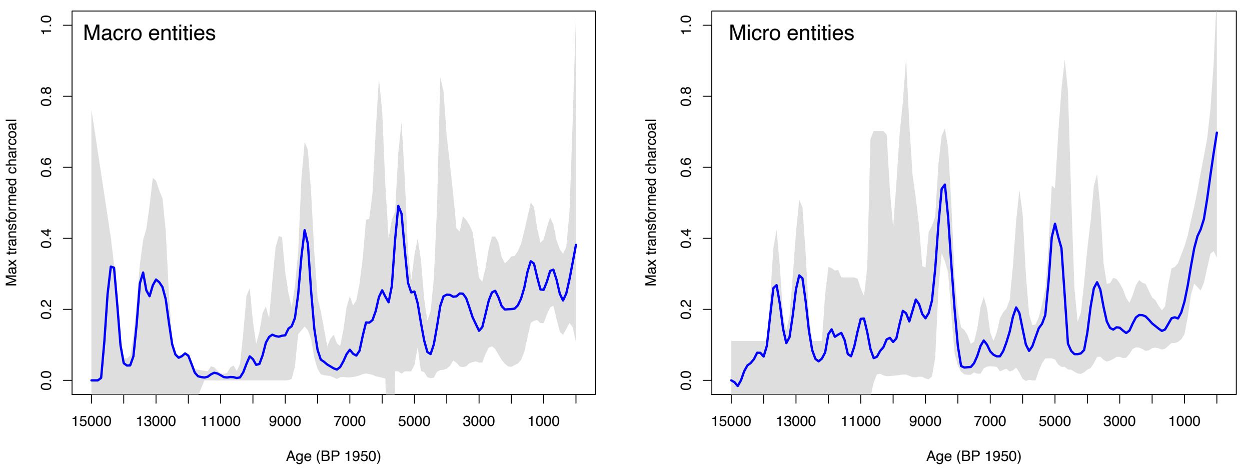
thus, ROS is proportional to the square of the fire duration.

However, the area covered by a fire is approximately proportional to the square of the distance covered by the fire front. Therefore, the area burnt is proportional to the fourth power of the duration of the fire. Any environmental factors influencing fire duration will be strongly amplified in burnt area, using the fourth root removes this amplification.

S4. Impact of using of micro and macro charcoal on the burnt area fraction reconstructions

The charcoal data used in these analyses was generated in several different ways and includes counts of both microcharcoal and macrocharcoal. Some sites had only macroscopic charcoal, some only microscopic charcoal, and a small number of sites included both. Macrocharcoal is often thought to be associated with local fires, and microcharcoal to represent regional fires. In order to test whether size had any impact of the composite reconstructions, we compared the 15 charcoal sites with both macroscopic and microscopic records (Fig. S3). Since this analysis suggests there is no difference in the curves obtained, we used both types of record in our analyses, though preferring macrocharcoal at those sites with both kinds of records.

Figure S3. Composite plots for macroscopic and macroscopic charcoal separately of 15 charcoal sites with both macroscopic and microscopic entities using the locfit() function with half-width=300, number of bootstrap samples=1000. The locally estimated scatterplot smoothing is shown in blue; The upper and lower 95th-percentile confidence intervals are shown in grey.



S5. Comparisons of WA-PLS and TWA-PLS reconstructions with and without fx correction

Weighted Averaging Partial Least-Squares (WA-PLS) regression (ter Braak et al., 1993; ter Braak and Juggins, 1993; Salonen et al., 2012) is widely used for climate reconstructions, but there is a known tendency for the reconstructed values to be compressed towards the middle of the range of the climate variable as expressed in the training data set. Tolerance-weighted Weighted Averaging Partial Least-Squares with a sampling frequency correction (fxTWA-PLS: Liu et al., 2020) is a modification of WA-PLS, designed to reduce the compression of reconstructions towards the centre of the climatic range sampled by the training dataset by accounting for the climatic tolerances of individual pollen taxa and the frequency (fx) of the sampled climate variable in the training dataset. Since fxTWA-PLS has not previously been used to reconstruct burnt area fractions, we tested whether this approach reduced compression in the burnt area fraction reconstructions when compared to

WA-PLS. We also tested the impact of using the sampling frequency correction (fx) separately for both WA-PLS and TWA-PLS. Cross-validation fitness assessment (Table S4) and visual comparison of the fitted plots and residuals (Fig. S4) indicate that fxTWA-PLS reduces the compression bias more than other methods and also has higher predictive power. The composite curves produced using each approach are shown in Fig. S5.

Table S4. Leave-out cross-validation fitness of WA-PLS and TWA-PLS methods, with and without fx correction, showing results for all the components. The last significant number of components are shown in bold.

Method	ncomp	R^2	RMSEP	ΔRMSEP	p	b_0	b_1	b_0.se	b_1.se
WA-PLS	1	0.292	0.049	-15.716	0.001	0.090	0.309	0.002	0.015
	2	0.373	0.046	-5.622	0.001	0.077	0.406	0.002	0.017
	3	0.422	0.044	-4.226	0.001	0.072	0.446	0.002	0.016
	4	**0.437**	**0.044**	**-1.196**	**0.003**	**0.069**	**0.465**	**0.002**	**0.017**
	5	0.453	0.043	-1.276	0.133	0.066	0.491	0.002	0.017
	6	0.461	0.043	-0.723	0.204	0.064	0.500	0.002	0.017
	7	0.463	0.043	-0.324	0.375	0.065	0.502	0.002	0.017
	8	0.469	0.043	-0.358	0.297	0.063	0.514	0.003	0.017
TWA-PLS	1	0.277	0.049	-14.917	0.001	0.097	0.269	0.002	0.014
	2	0.363	0.047	-5.683	0.001	0.077	0.399	0.002	0.017
	3	0.435	0.044	-5.952	0.001	0.069	0.465	0.002	0.017
	4	0.469	0.042	-3.101	0.001	0.065	0.498	0.002	0.017
	5	0.484	0.042	-1.524	0.019	0.063	0.510	0.002	0.017
	6	**0.497**	**0.041**	**-1.304**	**0.037**	**0.062**	**0.521**	**0.002**	**0.016**
	7	0.501	0.041	-0.313	0.360	0.061	0.528	0.002	0.017
	8	0.507	0.041	-0.554	0.201	0.060	0.539	0.002	0.017
WA-PLS with *fx* correction	1	0.291	0.055	-4.861	0.056	0.075	0.510	0.004	0.025
	2	0.363	0.053	-4.749	0.006	0.052	0.622	0.004	0.026
	3	0.415	0.048	-8.864	0.001	0.054	0.615	0.003	0.023
	4	0.431	0.047	-2.016	0.011	0.060	0.606	0.003	0.022
	5	**0.439**	**0.047**	**-0.669**	**0.008**	**0.059**	**0.614**	**0.003**	**0.022**
	6	0.420	0.048	1.896	0.959	0.066	0.581	0.003	0.021
	7	0.388	0.050	5.155	1.000	0.070	0.570	0.003	0.023
	8	0.409	0.049	-2.462	0.001	0.068	0.580	0.003	0.022
TWA-PLS with *fx* correction	1	0.276	0.054	-6.820	0.003	0.086	0.444	0.003	0.023
	2	0.365	0.050	-7.599	0.001	0.065	0.554	0.003	0.023
	3	0.435	0.047	-5.311	0.009	0.057	0.628	0.003	0.022
	4	**0.467**	**0.045**	**-4.852**	**0.001**	**0.061**	**0.610**	**0.003**	**0.020**
	5	0.454	0.046	1.615	0.849	0.062	0.603	0.003	0.021
	6	0.468	0.045	-1.735	0.005	0.062	0.606	0.003	0.020
	7	0.487	0.044	-1.455	0.094	0.057	0.638	0.003	0.021
	8	0.458	0.046	4.292	0.998	0.058	0.634	0.003	0.022

Figure S4. The fitted plots and residual plots of WA-PLS and TWA-PLS methods, with and without fx correction.

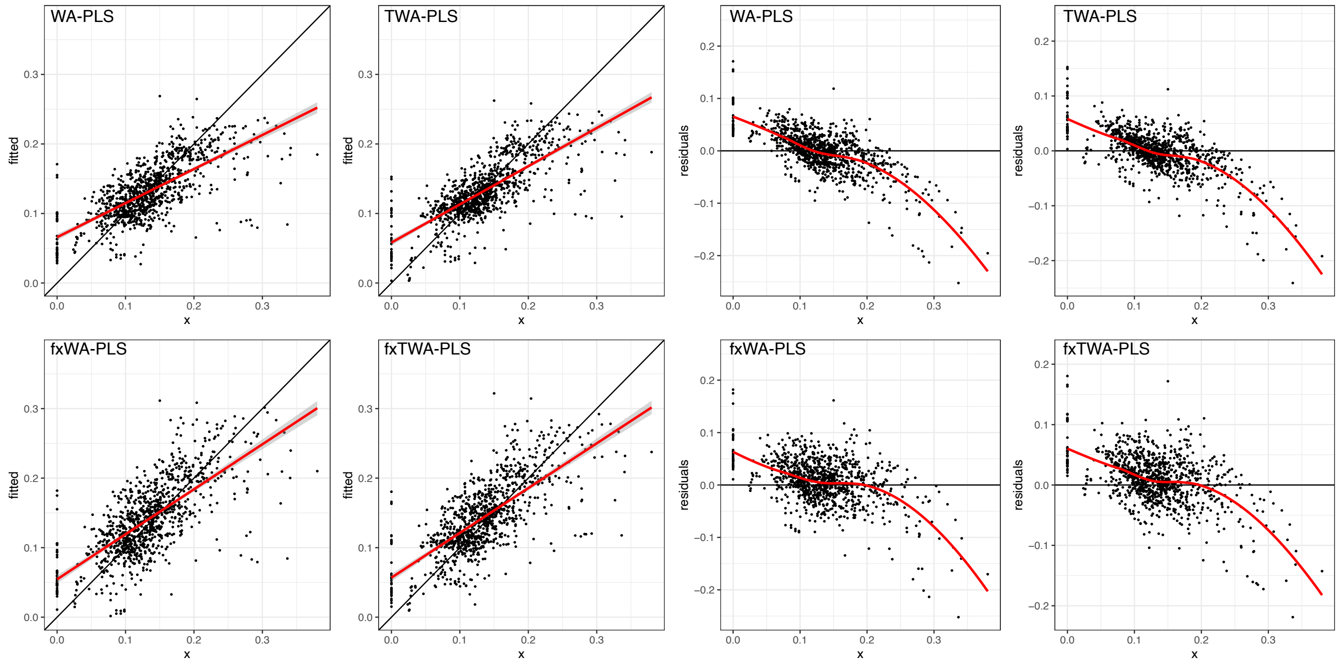
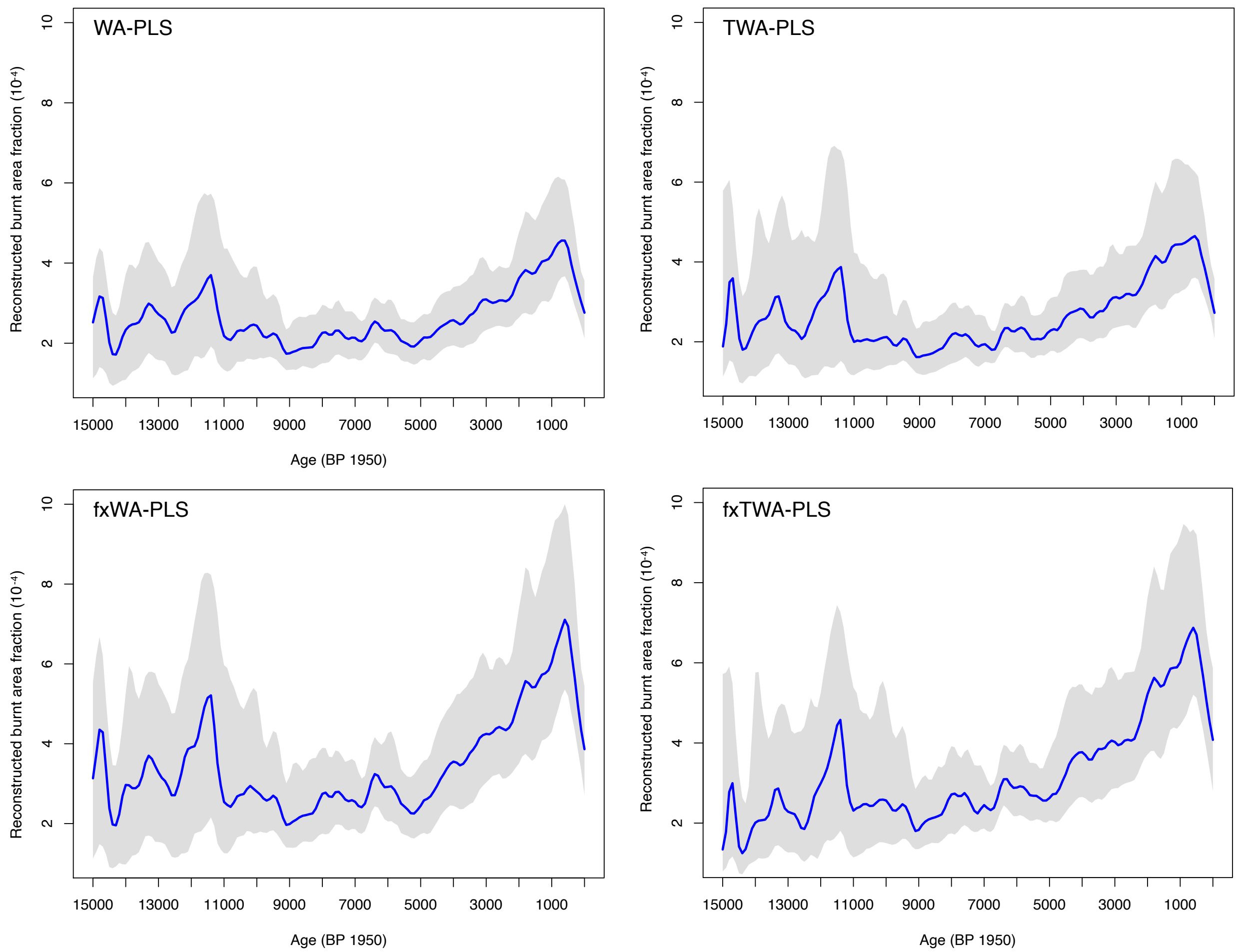


Figure S5. Composite curves of reconstructed burnt area using WA-PLS and TWA-PLS, with and without fx correction, using the locfit() function with half-width = 300, number of bootstrap samples = 1000. The locally estimated scatterplot smoothing is shown in blue; The upper and lower 95th-percentile confidence intervals are shown in grey.



References

ter Braak, C. J. F. and Juggins, S.: Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages, Hydrobiologia, 269–270(1), 485–502, doi:10.1007/BF00028046, 1993.

ter Braak, C. J. F., Juggins, S., Birks, H. J. B. and van der Voet, H.: Weighted averaging partial least squares regression (WA-PLS): Definition and comparison with other methods for species-environment calibration, in Multivariate Environmental statistics, edited by G. P. Patil and C. R. Rao, pp. 525–560, Elsevier, Amsterdam. [online] Available from: https://edepot.wur.nl/249353, 1993.

Hansen, M. and Song, X. P.: MEaSUREs Vegetation Continuous Fields (VCF) Yearly Global 0.05 Deg V001., NASA, L. Process. Distrib. Act. Arch. Cent., doi:10.5067/MEaSUREs/VCF/VCF5KYR.001, 2018.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O'sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U. and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17(5), 1343–1365, doi:10.5194/bg-17-1343-2020, 2020.

Klein Goldewijk, K., Beusen, A., Doelman, J. and Stehfest, E.: New anthropogenic land use estimates for the Holocene : HYDE 3.2, Earth Syst. Sci. Data, 9(2), 927–953, doi:https://doi.org/10.5194/essd-2016-58, 2017.

Liu, M., Prentice, I. C., Ter Braak, C. J. F. and Harrison, S. P.: An improved statistical approach for reconstructing past climates from biotic assemblages: Improved palaeoclimate reconstruction, Proc. R. Soc. A Math. Phys. Eng. Sci., 476(2243), doi:10.1098/rspa.2020.0346, 2020.

Salonen, J. S., Ilvonen, L., Seppä, H., Holmström, L., Telford, R. J., Gaidamavičius, A., Stančikaite, M. and Subetto, D.: Comparing different calibration methods (WA/WA-PLS regression and Bayesian modelling) and different-sized calibration sets in pollen-based quantitative climate reconstruction, Holocene, 22(4), 413–424, doi:10.1177/0959683611425548, 2012.

Viovy, N.: CRUNCEP Version 7 - Atmospheric Forcing Data for the Community Land Model, Res. Data Arch. Natl. Cent. Atmos. Res. Comput. Inf. Syst. Lab., doi:10.5065/PZ8F-F017, 2018.