Supplement of

Palynological evidence reveals an arid early Holocene for the north-east Tibetan Plateau

Nannan Wang et al.

Correspondence to: Xianyong Cao (xcao@itpcas.ac.cn)

Quantitative climate reconstruction of Gahai Lake

1. Modern pollen dataset and meteorological data

The modern pollen dataset (*n*=731) is derived from the eastern Tibetan Plateau, which spans from 94.07–103.02°E and 29.13–38.48°N with an elevation between 2515 and 5008 m a.s.l. (Fig.1 of the main text). This modern pollen dataset is derived from Cao et al. (2014) plus the recently published pollen data from Cao et al. (2021) and Wang et al. (2022).

Modern climate data were obtained from the Chinese Meteorological Forcing Dataset (CMFD; gridded near-surface meteorological dataset) which contains remote-sensing products, reanalysis datasets, and in situ station data between 1979 and 2018 (He et al., 2020). This dataset has high spatial and temporal resolution, and its high reliability has already been confirmed for the Tibetan Plateau. The climate data are assigned to the nearest 1 km × 1 km grid, which is calculated by smoothing spline interpolation of multi-annual means of climatic data from nearby meteorological stations, and this is achieved using the *rdist.earth* function in the *fields* package version 9.6.1 (Nychka et al., 2019) for R (version 3.6.0; R Core Team, 2020). We extracted climate data for each sample site of mean annual precipitation (P_{ann}), mean temperature of the coldest month (Mt_{co}) and warmest month (Mt_{wa}), and mean annual temperature (T_{ann}).

2. Establishment of the pollen-climate transfer function

Weighted averaging partial least squares regression (WA-PLS) was employed to evaluate the potential of the pollen dataset for past climate reconstruction and its performance was tested using "leave-one-out" cross-validation (ter Braak & Juggins, 1993) with R^2 (coefficient of determination between observed and predicted values) and RMSEP (root mean square error of prediction; Birks, 1998). The number of WA-PLS components used was selected using a randomisation *t*-test (Juggins and Birks, 2012). The climate reconstruction was completed using R software with the *rioja* package version 0.7–3 (Juggins, 2012), and the pollen assemblages were square-root transformed before reconstruction to reduce noise (Prentice, 1980).

3. Reliability of the pollen-climate transfer function and reconstructions for Gahai Lake

Ordination analysis indicated that P_{ann} and Mt_{wa} are important climatic determinants of pollen distribution, thus pollen–climate calibration-sets including P_{ann} and Mt_{wa} were established to assess the predictive power of this pollen dataset. The results of "leave-one-out" cross-validation showed that the first component for P_{ann} (R^2=0.61, RMSEP=109.58 mm) and Mt_{wa} (R^2=0.37, RMSEP=2.56 °C) performed well (Table S1).

Table S1. Model performance statistics as assessed by "leave-one-out" cross-validation for the five components of the weighted averaging partial least square regression (WA-PLS). RMSEP: root mean squared error of prediction; R^2: coefficient of determination between bootstrap predicted and observed values; Ave. Bias: the average bias of the parameter estimate; Max. Bias: the maximum bias of the parameter estimate; Rand. t-test: randomised t-test.

Variables	Method	RMSEP	R^2	Ave. Bias	Max. Bias	Rand. t-test
P_{ann}	**Component 1**	**109.58**	**0.6105**	**-0.60344**	**165.7842**	**0.001**
	Component 2	106.33	0.6367	-0.17538	173.9644	0.048
	Component 3	104.33	0.6493	-0.08673	150.1874	0.077
	Component 4	104.37	0.6501	-1.22009	149.6598	0.527
	Component 5	104.88	0.6479	-0.02029	155.7057	0.800
Mt_{wa}	**Component 1**	**2.55976**	**0.36841**	**0.05264**	**6.02357**	**0.001**
	Component 2	2.49063	0.40540	0.07848	5.89699	0.025

Component 3	2.49095	0.40748	-0.00082	5.39354	0.526
Component 4	2.51670	0.40148	0.01266	4.97078	0.872
Component 5	2.55368	0.39158	-0.00821	5.04967	0.935

We argue in detail that the arboreal pollen should be considered as exogenous components before 7.4 ka BP in the main text (Discussion 5.2), and we reconstructed the climate based on the fossil pollen record either including or excluding arboreal pollen taxa to investigate the potential ranges of P_{ann} and Mt_{wa} between 14.2 and 7.4 ka BP. Including arboreal pollen in this time period (darker line) reconstructs more precipitation than excluding arboreal pollen (lighter line). After 7.4 ka BP, the pollen data of dark and light curves are the same, hence we believe that the reconstruction including arboreal pollen (dark curve) is unbiased and it is thus unnecessary to show the whole light curve (excluding arboreal taxa). We infer from this that arboreal pollen of the last 7.4 ka in the pollen assemblages mainly comes from within the catchment,

The quantitative reconstructions show that the exogenous arboreal pollen taxa have no significant effect on Mt_{wa}, but do have a great impact on P_{ann} between 14.2 and 7.4 ka BP (Fig. S1). From 14.2 to 7.4 ka BP, P_{ann} ranges from 400 to 734 mm, and Mt_{wa} varies between 7.7 and 10°C. Mt_{wa} shows a slight decrease during 10.8–7.4 ka BP, whereas P_{ann} has no significant change compared with the former stage. The highest values of P_{ann} and Mt_{wa} occur between 7.4 and 3.8 ka BP. After 3.8 ka BP, P_{ann} holds low values with little temporal changes while Mt_{wa} decreases continuously (Fig. S1).

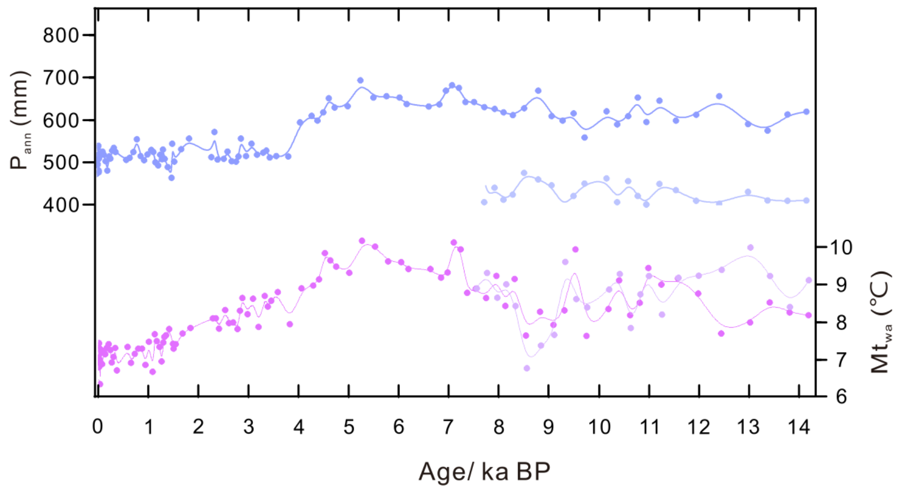

Figure S1. Reconstructions of P_{ann} and Mt_{wa} for Gahai Lake based on the fossil pollen record either including (darker lines) or excluding (lighter lines) arboreal pollen taxa before 7.4 ka BP.

Supplementary References

Birks, H. J. B.: Numerical tools in quantitative palaeolimnology: Progress, potentialities, and problems, J. Paleolimnol., 20, 301–332, https://doi.org/10.1023/A:1008038808690, 1998.

Cao, X., Herzschuh, U., Telford, R. J., and Ni, J.: A modern pollen-climate dataset from China and Mongolia: Assessing its potential for climate reconstruction, Rev. Palaeobot. Palynol., 211, 87–96, https://doi.org/10.1016/j.revpalbo.2014.08.007, 2014.

Cao, X., Tian, F., Li, K., Ni, J., Yu, X., Liu, L., and Wang, N.: Lake surface-sediment pollen dataset for the alpine meadow vegetation type from the eastern Tibetan Plateau and its potential in past climate reconstructions, Earth Syst. Sci. Data, 13, 3525–3537, https://doi.org/10.5194/essd-13-3525-2021, 2021.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Sci. Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020.

Juggins, S.: Rioja: analysis of Quaternary Science Data version 0.7-3, available at: http://cran.r-project.org/web/packages/rioja/index.html (last access: June 2020), 2012.

Juggins, S., and Birks, H. J. B.: Quantitative environmental reconstructions from biological data, in: Birks, H. J. B., Lotter, A. F., Juggins, S., and Smol, J. P., Tracking environmental change using lake sediments (vol. 5): Data handling and numerical techniques, Springer, Dordrecht, 431–494, 2012.

Nychka, D., Furrer, R., Paige, J., and Sain, S.: *fields: Tools for spatial data, version 9.6.1*, available at: https://cran.r-project.org/web/packages/fields/ (last access: June 2020), 2019.

R Core Team.: *R, A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, 2020.

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, 485–502, doi: 10.1007/BF00028046, 1993.

Wang, N., Liu, L., Zhang, Y., and Cao, X: A modern pollen dataset for the forest-meadow-steppe ecotone from the Tibetan Plateau and its potential use in past vegetation reconstruction, Boreas, https://doi.org/10.1111/bor.12589, 2022.