A lacustrine surface-sediment pollen dataset covering the Tibetan Plateau and its potential in past vegetation and climate reconstructions

Fang Tian[1], Weiyu Cao[1], Xiaohan Liu[1], Zixin Liu[1], Xianyong Cao[2]

[1] *College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China*

[2] *Group of Alpine Paleoecology and Human Adaptation (ALPHA), State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources (TPESER), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China*

Correspondence: Fang Tian (tianfang@cnu.edu.cn)

Abstract. A dataset of pollen extracted from the surface-sediments of lakes with broad spatial coverage is essential for pollen-based reconstructions of past vegetation and climate. We collected 90 new lake surface-sediment pollen samples from the Tibetan Plateau (TP), covering major vegetation types, including alpine forest, alpine meadow, alpine steppe, and alpine desert. By integrating these new data with previously published lacustrine pollen datasets, we established a comprehensive modern pollen dataset comprising 476 samples across the TP, covering the full range of climatic gradients across the TP, with net primary production (NPP) from 0.16 to 6617.36 Kg C m^{-2}, mean annual precipitation (P_{ann}) from 97 to 788 mm, mean annual temperature (T_{ann}) -9.09 to 6.93 °C, mean temperature of the coldest month (Mt_{co}) -23.48 to -2.65°C, and mean temperature of the warmest month (Mt_{wa}) 1.77 to 19.26°C. Numerical analyses based on the comprehensive modern pollen dataset (n=476) revealed that P_{ann} is the primary climatic determinant for pollen distribution, while NPP is a valuable variable reflecting vegetation conditions. To detect the quantitative relationship between pollen and NPP/P_{ann}, both weighted-averaging partial least squares (WA-PLS) and random forest algorithm (RF) were employed. The performance of both models suggests that this modern pollen dataset has good predictive power in estimating past NPP and P_{ann}, but RF has a slight advantage with this dataset. This comprehensive modern pollen dataset provides a reliable basis for reconstructing past vegetation and climate changes on the central TP. However, caution is required when applying it to pollen spectra from marginal regions of the TP or to records covering the Last Glacial period, where analogue quality is relatively poor. The dataset, including site locations, pollen percentages, NPP, and climate data for 90 lakes, is available at the National Tibetan Plateau Data Center (TPDC; Tian et al., 2025; https://doi.org/10.11888/Paleoenv.tpdc.302470).

删除了: an even

删除了: distribution

删除了:)

删除了: . A comprehensive modern pollen dataset is established by

删除了: our newly obtained modern pollen dataset

删除了: previous modern

删除了: net primary production (

删除了:)

删除了: /NPP

删除了: is considered

删除了: when

删除了: from pollen spectra from

删除了: , but

删除了: needed if

删除了: is applied

删除了: the

删除了: and those

删除了: due to poor

删除了: in those cases

删除了: ,

1 Introduction

A modern pollen dataset is the foundation for the quantitive reconstruction of past vegetation and climate based on fossil pollen spectra. Surface-soil samples for pollen analysis can be easily obtained, but their pollen assemblages are easily affected by local vegetation components, which cause more noise in the modern relationships of pollen–vegetation and pollen–climate (Cao et al., 2014). Sediment from lakes, in contrast, provide more regional pollen signals owing to broader pollen source areas, more stable sedimentation rates, and better preservation, making them more suitable for regional vegetation and climate changes (Tian et al., 2020; Cao et al., 2021). Due to the sparse distribution of lakes, high sampling costs, and limited accessibility—especially in remote regions—modern pollen datasets from lake surface sediments remain limited and spatially biased, particularly in China (Herzschuh et al., 2010; Ma et al., 2017; Cao et al., 2021).

The Tibetan Plateau (TP), situated at high elevations and subject to complex climate systems, is highly sensitive to global climate change and human activities and exhibits strong regional ecological and climatic peculiarities (Chen et al., 2015, 2020; Pepin et al. 2019). These features make the TP a research hotspot for past vegetation and climate reconstructions. Fortunately, the widespread distribution of lakes across the plateau offers an opportunity to expand and refine pollen-based calibration datasets using lake surface sediments, but the distribution of available pollen sites of lake surface-sediment remains uneven and incomplete due to logistical constraints (Cao et al., 2021; Qin, 2021; Ma et al., 2024). Hence, it is essential to improve the coverage and comprehensiveness of the modern calibration-set from lake surface-sediments on the TP.

While previous pollen–climate relationships are often the focus of calibration-set studies, the pollen–vegetation relationship is also crucial on the TP, where vegetation type is generally employed as the target variable, especially when reconstructing ecological conditions (e.g. Qin, 2021; Qin et al., 2022). Existing modern pollen datasets reveal that pollen assemblages from different vegetation types on the TP generally present only minor differences in pollen components and their abundance. For instance, the dominant pollen taxa are generally herbaceous taxa, including Cyperaceae, *Artemisia*, Amaranthaceae (synonym: Chenopodiaceae), and Poaceae (e.g. Herzschuh et al., 2010; Ma et al., 2017; Cao et al., 2014, 2021; Li et al., 2020; Qin, 2021), making it difficult to distinguish vegetation conditions based on pollen assemblages directly. However, the pollen concentration and percentages from lake sediments have been confirmed to positively correlate with vegetation coverage, which reflects total plant biomass (Liu et al., 2023). Since net primary production (NPP) represents the carbon fixed and accumulated as biomass by plants (Fang et al., 2001; Nemani et al., 2003;

删除了: ↵

删除了: climate

删除了: vegetation

删除了: Situated

删除了: the Tibetan Plateau (TP)

删除了: climate and

删除了: .,

删除了: datasets

删除了: Previous

删除了: while

删除了: ,

删除了: However, the available

删除了: (=

删除了: Net

删除了:), which quantifies the amount of atmospheric

删除了: by plants

删除了: , plays an important role in the global carbon cycle

Gonsamo et al., 2013; Ni, 2013; Walker et al., 2015; Ji et al., 2020). pollen can serve as an indirect proxy for NPP, allowing us to infer spatial and temporal patterns of vegetation conditions on the TP.

Here, we analysed 90 lake surface-sediment samples for pollen and combined them with previously published 386 modern pollen data extracted from lake surface-sediments (Herzschuh et al., 2010; Li and Li, 2015; Cao et al., 2021; Ma et al., 2024; Wu et al., 2024), then used a combination of ordination techniques, weighted averaging partial least squares (WA-PLS), and Random Forest (RF) to: (1) establish a comprehensive pollen dataset extracted from lake surface-sediments covering the entire TP with an relative even distribution; (2) evaluate the predictive power of models using the modern pollen dataset in reconstructing past vegetation and climate.

2 Study area

The climate of the TP is controlled mainly by the Asian Summer Monsoon in summer with warm-wet conditions and by westerlies in winter with a cold-dry climate (Wang, 2006). In addition, there is a gradient from high summer temperatures (up to 19°C) and high precipitation (>700 mm) on the south-eastern TP, to low summer temperatures (ca. 6°C) and low precipitation (<100 mm) on the north-western TP (Fig. 1; Sun, 1999; Herzschuh, 2007; He et al., 2020).

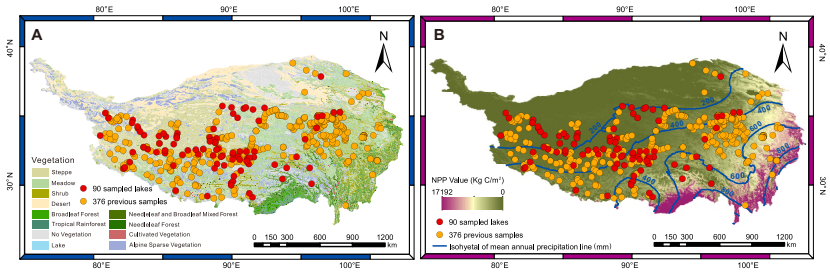

Figure 1. Spatial distribution of 476 modern pollen samples collected from lake surface-sediments on the Tibetan Plateau (red filled circles: 90 sampled lakes; orange filled circles: 386 previous samples; Herzschuh et al., 2010; Li and Li, 2015; Cao et al., 2021; Ma et al., 2024; Wu et al., 2024) based on (A) vegetation types and (B) net primary production (NPP, Zhao and Running, 2010).

The TP exhibits distinct vegetation zonation along its south-east–north-west thermal and moisture gradients, progressing from forest ecosystems through alpine meadows and steppes to desert vegetation (Fig. 1; Zhang, 2007). Alpine forest dominated by *Pinus*, *Picea*, *Abies*, *Betula*, *Quercus*, and *Tsuga* are primarily distributed in the warm-humid south-eastern and eastern marginal regions of the TP (Herzschuh, 2007). Alpine meadow, as one of the most important vegetation types, are mainly distributed on the eastern and southern TP, and are characterized by *Kobresia* spp., *Carex*, Asteraceae,

删除了:). Therefore, NPP may
删除了: a more sensitive alternative variable in reflecting the
删除了: distribution
删除了: change
删除了: climate and
删除了: C
删除了: ,
删除了: F
删除了: s
删除了: s

Polygonum, *Potentilla*, Fabaceae, Caryophyllaceae, *Leontopodium*, *Arenaria*, *Ranunculus,* and Poaceae (Wu, 1995; Herzschuh et al., 2010; Cao et al., 2021). Alpine steppe are primarily distributed across the southern, eastern, and central TP, and is mainly dominated by *Stipa purpurea*, *Artemisia*, *Potentilla*, Asteraceae, Amaranthaceae, and *Carex* (Fig. 1; Zhang, 2007; Yue et al., 2011). Alpine desert, located in the dry north-central and westernmost central TP, are characterized by sparse vegetation and are predominantly occupied by drought-tolerant taxa such as *Ceratoides* (Amaranthaceae), *Salsola*, *Haloxylon*, *Kalidium*, *Artemisia*, *Ephedra*, *Nitraria,* and Poaceae (Fig. 1; Zhang, 2007).

3 Materials and methods

3.1 Sample collection and pollen processing

To achieve a broadly representative coverage of lakes across different vegetation zones on the TP, we collected one surface-sediment sample (top 2 cm) from the centre of each lake, for a total of 90 lakes across different vegetation types on the TP: forest (n=5), meadow (n=22), steppe (n=53), and desert (n=10) between 2021 and 2023 (Fig. 1, Table 1). Collecting from the lake centre is intended to provide a representative pollen assemblage that integrates inputs from the surrounding catchment. The elevation of the sampled lakes ranges from 3923 to 5433 m a.s.l., with a median of 4652 m a.s.l. (Fig. 1).

Table 1. Locations of the sampling sites of our field work on the Tibetan Plateau.

No.	Lake	Latitude (°N)	Longitude (°E)	Elevation (m a.s.l.)	Water area (km^2)	Vegetation type
1	Cuomujiri	94.4304	29.8118	4235	1.64 [a]	Forest
2	Ranwu Lake	96.8252	29.3962	5263	13.58 [b]	Forest
3	Sanse Lake	94.7670	30.7239	4042	0.63 [a]	Forest
4	Ren Co	96.6748	30.7156	4452	3.55 [b]	Forest
5	Potal Lake	95.5743	31.6223	4656	8.35 [a]	Forest
6	Ruba Lake	90.1725	29.4644	3923	0.67 [b]	Meadow
7	Namucoluo	90.3347	29.6070	4690	0.36 [a]	Meadow
8	Cuoriwang	90.4064	30.0345	4400	0.28 [a]	Meadow
9	Niangde Co	90.1834	29.2810	4365	0.27 [a]	Meadow
10	Cona Lake	91.4305	32.0779	4602	188.54 [b]	Meadow
11	Tangbin Lake	90.9672	30.4795	5025	0.72 [a]	Meadow
12	Cuoe	91.5350	31.5088	4511	86.62 [b]	Meadow
13	Changma Lake	92.1069	32.0639	4932	3.89 [a]	Meadow
14	Cuomuri	92.0596	31.6201	4547	3.65 [a]	Meadow
15	Gemu Co	91.6990	31.5550	4524	1.43 [b]	Meadow
16	Xiongmu Co	91.6303	31.0399	4662	2.35 [a]	Meadow
17	Nairi Pingco	91.4788	31.2730	4513	95.56 [b]	Meadow
18	Cuomuzhelin	88.2168	28.3933	4395	54.43 [b]	Meadow

删除了: s
删除了: s
删除了: ,
删除了: ensure the even distribution
删除了: the sampled lakes
删除了: lake
删除了: samples (the
删除了: , n=90
删除了: in
删除了: vegetation types on the TP
删除了: range
删除了: these
删除了: varies
删除了: .

19	Nariyong Co	91.9377	28.3071	4731	23.31 [b]	Meadow
20	Peiku Co	85.5869	28.8507	4561	272.95 [b]	Meadow
21	Zhegu Co	91.6770	28.6316	4601	59.06 [b]	Meadow
22	Nianjie Co	96.2905	33.0773	4441	20.66 [b]	Meadow
23	Samu Co	93.7813	30.9753	4748	1.64 [a]	Meadow
24	Hala Lake	97.5967	38.2507	4071	611.57 [a]	Meadow
25	Zhaling Lake	97.3420	34.9447	4280	526.62 [b]	Meadow
26	Koucha Lake	97.2311	34.0081	4518	18.16 [b]	Meadow
27	Eling Lake	97.7130	35.0217	4257	629.75	Meadow
28	Goulu Co	92.4546	34.5942	4639	25.89 [a]	Steppe
29	UlanUl Lake	90.7108	34.8528	4857	566.96 [b]	Steppe
30	Xijir Ulan Lake	90.3528	35.1875	4769	373.87 [b]	Steppe
31	Lexiewudan Lake	90.2053	35.7071	4862	247.58 [b]	Steppe
32	Xiangyang Lake	89.4616	35.8194	4843	3.67 [b]	Steppe
33	Kekexili Lake	91.2205	35.6115	4875	315.95 [b]	Steppe
34	Kekao Lake	91.3874	35.6973	4881	60.04 [b]	Steppe
35	Zhuonai Lake	91.9833	35.5325	4734	255.37 [a]	Steppe
36	Kusai Lake	92.9412	35.6753	4471	271.08 [b]	Steppe
37	Zigêtang Co	90.8973	32.0674	4538	225.55 [b]	Steppe
38	Daru Co	90.7324	31.6562	4675	134.27 [b]	Steppe
39	Bange Lake	89.4734	31.7282	4519	136.34 [b]	Steppe
40	Lingge Co	88.7220	33.9370	5061	108.14 [b]	Steppe
41	Qiagang Co	88.3966	33.2313	4719	21.64 [a]	Steppe
42	Caiduochaka Lake	88.9793	33.1576	4833	37.14 [b]	Steppe
43	Eya Co	88.6713	33.0013	4824	75.14 [b]	Steppe
44	Ri Co	89.6068	30.9302	4648	113.02 [a]	Steppe
45	Mujiu Co	89.0144	31.0337	4664	83.58 [b]	Steppe
46	Suo Co	90.9056	31.3978	4556	0.16 [a]	Steppe
47	Mading Co	90.2995	31.4147	4680	0.16 [a]	Steppe
48	Maiding Co	90.3202	31.8413	4773	0.35	Steppe
49	Changma Co	87.8756	32.2605	4725	4.43 [b]	Steppe
50	Cuolongjiao	88.8539	32.7857	4873	3.55 [b]	Steppe
51	Duomaxiang Lake	89.1268	32.3249	4704	0.07 [a]	Steppe
52	Gewa Co	88.7968	30.6725	4745	1.86 [a]	Steppe
53	Wojiong Co	89.3646	31.6276	4598	0.01 [a]	Steppe
54	Gaa Co	88.9583	32.2130	4602	10.28 [b]	Steppe
55	Chelachapuka	86.1548	31.8024	4773	0.37 [a]	Steppe
56	Yong Co	84.7044	31.9383	4712	2.09 [b]	Steppe
57	Rena Co	84.2559	32.7281	4579	19.49 [b]	Steppe
58	Chabo Co	84.2108	33.3512	4500	40.89 [b]	Steppe
59	Jibuchaga Co	83.9975	32.0205	4467	8.17 [a]	Steppe
60	Cuoguo Co	83.2921	32.2503	4669	10.07 [a]	Steppe
61	Bieruoze Co	82.9417	32.4308	4392	32.27 [b]	Steppe
62	Shekazhi	82.0466	32.0115	4591	17.78 [a]	Steppe

删除了: Haling

删除了: Gelu

63	Dagze Co	87.4456	31.8332	4465	269.07 [b]	Steppe
64	Xiabie Co	87.2680	32.2179	4592	17.85 [b]	Steppe
65	Jiaruo Co	86.6001	32.1730	4445	13.34 [b]	Steppe
66	Xuguo Co	90.3251	31.9542	4598	33.12 [b]	Steppe
67	Beilei Co	88.4296	32.9120	4797	25.23 [b]	Steppe
68	Unknown	81.7962	31.1937	5433	0.50 [a]	Steppe
69	Nading Co	85.4359	32.6776	4845	12.43 [a]	Steppe
70	Bala Co	82.9849	33.4281	4757	1.53 [a]	Steppe
71	Dong Co	84.7120	32.1440	4388	92.47 [b]	Steppe
72	Xiaogemu Co	85.7384	33.5778	4711	0.04 [a]	Steppe
73	Ningri Co	85.6752	33.3333	5020	21.02 [b]	Steppe
74	Guping Lake	85.6787	33.1683	5030	2.22 [b]	Steppe
75	Qiuruba Co	84.7966	33.3073	4733	9.02 [b]	Steppe
76	Caima'er Co	84.5879	33.5469	4573	45.12 [b]	Steppe
77	Selin Co	88.6979	31.7363	4512	2129.02 [b]	Steppe
78	Zhari Namco	85.4004	30.9068	4595	990.26 [b]	Steppe
79	Kuhai Lake	99.1636	35.3070	4117	48.50 [b]	Steppe
80	Donggi Cona	98.6596	35.2875	4066	238.15 [a]	Steppe
81	Aru Co	82.4768	33.9682	4904	91.22 [a]	Desert
82	Aksai Chin Lake	79.7863	35.2456	4831	170.22 [a]	Desert
83	Kunchuke Co	82.6590	33.7096	5042	22.90 [b]	Desert
84	Xiawei Lake	82.0454	34.6738	5110	5.52 [b]	Desert
85	Luotuo Lake	81.9849	34.4339	5082	63.43 [b]	Desert
86	Meima Co	82.4404	34.1278	4897	145.22 [b]	Desert
87	Lhanag Co	81.2820	30.6674	4577	270.32 [a]	Desert
88	Hongshan Lake	80.0545	34.8300	5043	6.35 [b]	Desert
89	Manasarovar Lake	81.3939	30.7465	4577	409.90 [b]	Desert
90	Xiada Co	79.3584	33.3916	4338	8.04 [a]	Desert

[a]: data measured from ArcGis; [b] data from Yang, 2019.

For each sample, 2–3 g of dry material was used for pollen extraction, and a tablet with *Lycopodium* spores (10,315 grains) was added to each sample initially as a tracer (Maher, 1981). Pollen samples were processed using standard acid–alkali–acid procedures (Fægri and Iversen, 1989), including 10% HCl, 10% KOH, 40% HF, acetolysis treatment, and sieving in an ultrasonic bath to remove particles <7 μm. Pollen grains were identified and counted under a Zeiss optical microscope at 400× magnification, referring to modern pollen slides collected from the eastern and central TP and published palynological literatures (Wang et al., 1995; Tang et al., 2016; Cao et al., 2020). To ensure the reliability of the pollen assemblages for numerical analyses, more than 500 terrestrial pollen grains, or over 2000 *Lycopodium* spores were counted for each sample. The pollen diagram was constructed using Tilia software (Grimm, 1987, 1991).

删除了: tracers

3.2 Data collection and harmonization

We compiled a dataset of modern pollen assemblages from lake surface sediments across the TP, incorporating 375 lakes situated in the eastern (Herzschuh et al., 2010; Cao et al., 2021), central, and western TP (Ma et al., 2024; Wu et al., 2024), obtained from accessible databases or from authors directly. To enhance spatial coverage, an additional 11 surface pollen assemblages were digitized from published diagram representing sites along the eastern edge of TP (Li and Li, 2015). The final dataset comprises 476 pollen assemblages from lake surface-sediments on the TP (Fig. S1). The pollen assemblages of the 386 previously published samples have already been described and discussed in detail in their original publications. Therefore, in this study, we present only the pollen assemblages of the 90 newly collected samples.

The pollen data are standardized following the procedures outlined in Cao et al. (2013), including harmonization of taxonomy generally to the family or genus level, and recalculation of pollen percentages based on total terrestrial pollen grains. Only pollen taxa with an abundance of at least 0.5% in at least three samples and a maximum ≥3% (*n*=35) were retained for the following statistical analyses (RDA, WA-PLS, and RF).

We employed the Chinese Meteorological Forcing Dataset (CMFD), a gridded near-surface meteorological dataset covering the period from January 1979 to December 2018, with a temporal resolution of 3 h and a spatial resolution of 0.1°. Climate data of each sampled lake were assigned as the values of the nearest pixel from the meteorological dataset. For all 476 lakes, the following parameters were extracted: P_{ann}: mean annual precipitation, mm; T_{ann}: mean annual temperature, °C; Mt_{co}: mean temperature of the coldest month, °C; Mt_{wa}: mean temperature of the warmest month, °C (He et al., 2020). The geographical distances between lake coordinates and grid centroids were calculated geodetically using the *rdist.earth* function in the *fields* package version 16.3.1 (Nychka et al., 2025) for R (R Core Team, 2019).

The NPP value, defined as Gross Primary Productivity (GPP) minus Maintenance Respiration (MR) (Zhao and Running, 2010), was obtained from observations of the MOD17A3HGF.006 product during 2001–2022 with a pixel resolution of 1000 m. Across the study region, NPP values range from 0.16 to 6617.36 Kg C m^{-2}, P_{ann} ranging from 97 to 788 mm, and cold thermal conditions characterized by low T_{ann} (-9.09 to 6.93°C) and Mt_{co} (-23.48 to -2.65°C; Table S1).

3.3 Data analysis

For all statistical analyses (redundancy analysis: RDA, weighted averaging partial least squares regression: WA-PLS, and Random Forest: RF), we used the full integrated dataset of 476 samples.

To visualize how the modern pollen assemblages respond to climatic variables, ordination techniques were employed based on the selected 35 pollen types from all 476 sites. Pollen data were square-root transformed to stabilize variances and optimize

删除了: Tibetan Plateau

删除了: total

删除了: A1).

删除了: –

删除了: –

删除了: *n*=35

删除了: each lake

删除了: the

删除了: ; Nychka et al., 2025)

删除了: 2

删除了: .

the signal-to-noise ratio (Prentice, 1980). Detrended correspondence analysis (DCA; Hill and Gauch, 1980) showed that the gradient length of the first axis of the pollen data was 2.36 SD (standard deviation units), indicating that a linear response model is suitable for our pollen dataset (ter Braak and Verdonschot, 1995). We employed RDA to assess how major pollen taxa and sampling sites are distributed along vegetation and climate gradients. Climatic predictors were introduced sequentially following a forward selection procedure, with multicollinearity assessed at each step via variance inflation factors (VIF). Variables exhibiting VIF values above the threshold of 20 were excluded to maintain model parsimony and reduce redundancy (ter Braak and Prentice, 1988; Birks, 1995). Additionally, the suitability of each climatic variable for quantitative reconstruction was evaluated using the ratio of the first constrained eigenvalue to the first unconstrained eigenvalue (λ_1/λ_2), where larger ratios indicate stronger predictive potential (Juggins, 2013). All ordinations were carried out using the *rda* and *decorana* functions in the *vegan* package (Oksanen et al., 2019).

WA-PLS regression was applied to calibrate transfer functions linking modern pollen assemblages to P_{ann} and NPP, based on square-root transformed relative abundances of the 35 selected taxa—consistent with those used in the ordination analyses (ter Braak and Juggins, 1993). Model performance was evaluated using "leave-one-out" cross-validation, and the optimal number of WA-PLS components was determined based on a randomization *t*-test (Juggins and Birks, 2012). All the analyses were performed using the *WA-PLS* function of the *rioja* package version 0.7–3 (Juggins, 2012) in R.

As WA-PLS is known to produce systematic prediction biases near the ends of environmental gradients—commonly referred to as the "edge effect" (Birks, 1998; Tian et al., 2022) —we further explored a complementary reconstruction method. RF is an ensemble learning algorithm that integrates multiple decision trees based on a classification tree algorithm and summarizes their results for classification or regression tasks (Breiman, 2001). The importance of the explanatory variable is normally measured as a percentage increase in the residual sum of squares after random shuffling of the variables order, thereby determining which explanatory variable can be added to the model. RF has been applied in the geographical and ecological fields and performs well (Li, 2013; Jin et al., 2016). In this study, we applied RF to assess the importance of pollen and the NPP/climate variables (Table S1). The model was systematically optimized through a stepwise reduction procedure, in which the pollen taxa with the least important score was deleted until the RF-importance of all remaining taxa were greater than 0 (Breiman, 2001). The RF algorithm was run based on square-root transformed pollen percentages of the selected 35 taxa, using the *randomForest* function in the *randomForest* package version 4.6–14 (Liaw, 2018) in R. The statistical significance of the reconstructions derived from WA-PLS and RF were tested with the *randomTF* function of the *palaeoSig* package (Telford and Birks, 2011; Telford, 2013)

删除了: redundancy analysis (

删除了:)

删除了: ere

删除了: Random Forest (

删除了:)

删除了: order of the

删除了: establish

in R.

In quantitative climate reconstructions, the taxonomic distance between a fossil pollen assemblage and its modern analogue is a key variable in evaluating the analogue quality (Birks et al., 1990). Shorter distances indicate closer taxonomic similarity and higher analogue quality, enhancing reconstruction reliability. This distance is commonly calculated using the squared chord distances based on the percentages of all pollen taxa. To evaluate the analogue quality, we calculated the squared chord distances between the selected fossil pollen spectra since the last glacial maximum (*n*=65, elevation higher than 3000 m a.s.l.; Cao et al., 2013) and the combined modern pollen dataset on the TP. The square chord distances were calculated using the *MAT* function of the *rioja* package (Juggins, 2018) in R.

4 Data description

The pollen assemblages of the new surface-sediment samples (*n*=90) are dominated by herbaceous pollen from alpine meadow, steppe, and desert sites on the TP. In contrast, arboreal pollen dominates the samples collected from forest, consisting mainly of *Pinus*, *Picea*, *Alnus*, *Tsuga*, *Juniperus*, *Betula*, and *Quercus* (Fig. 2). Additionally, there are evident regional peculiarities in its distribution (Figs. 2–4). Sites with Cyperaceae abundances >60% from alpine meadows are more common than other sites, whereas steppe regions are marked by higher percentages of Poaceae and *Artemisia*, typically exceeding 30% and 50%, respectively. The distribution center of Amaranthaceae (> 30%) is generally located in desert (Figs. 2–4; Table S1).

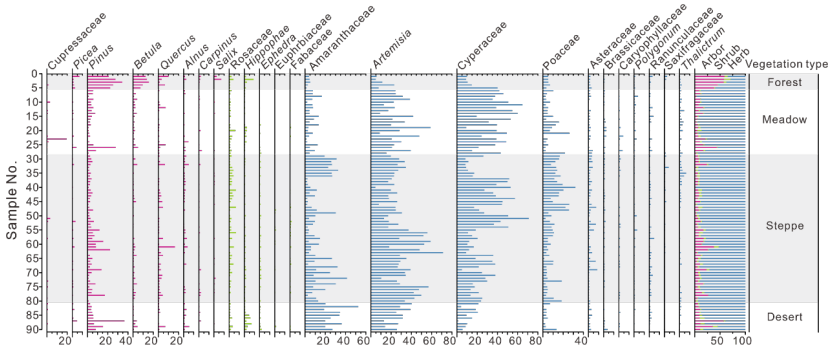

Figure 2. Percentage diagram of major pollen taxa for 90 lake surface-sediment samples on the Tibetan Plateau. Samples are arranged according to their vegetation type.

Group 1 (forest, *n*=5): The pollen assemblages of the sampled lakes are characterized by the lowest *Artemisia* and Amaranthaceae abundance, yet exhibits the highest arboreal pollen (AP) percentages among the four groups. Key arboreal taxa include

删除了: Fig
删除了: Fig
删除了: 2
删除了: content

Pinus (mean 26.0%, maximum 34.2%), *Betula* (mean 11.7%, maximum 15.6%), *Quercus* (mean 3.9%, maximum 9.3%), and *Picea* (mean 2.7%, maximum 7.0%, Figs. 2–4).

Group 2 (meadow, *n*=22): This group is typically characterized by the lowest AP and A/Cy (*Artemisia*/Cyperaceae) ratio but the highest Cyperaceae abundance (mean 39.8%, maximum 64.7%), with common taxa comprising *Artemesia* (mean 27.1%, maximum 58.9%), Amaranthaceae (mean 6.8%, maximum 16.4%), and Poaceae (mean 6.3%, maximum 26.1%, Figs. 2–4).

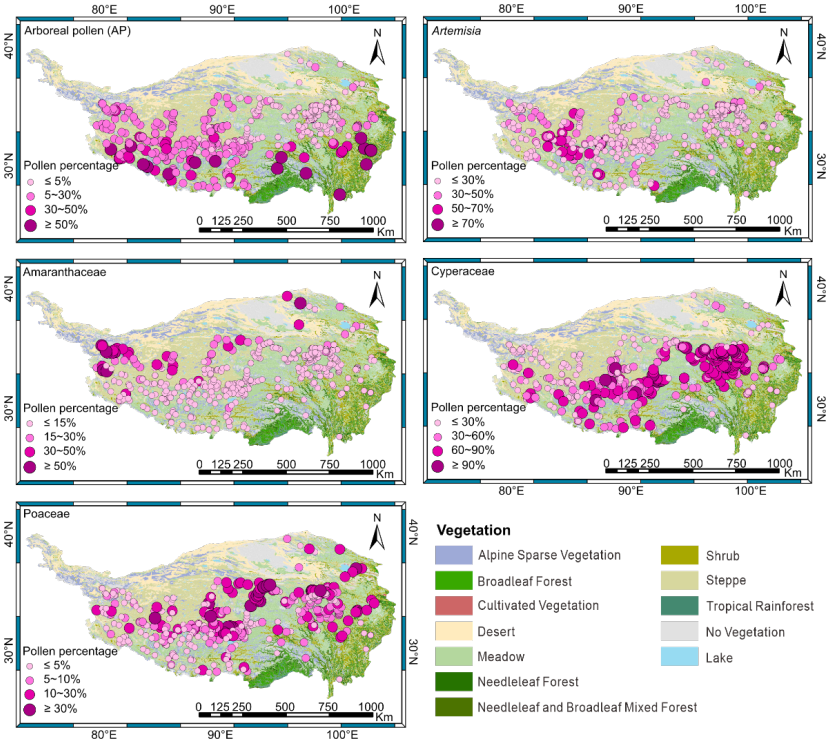

Figure 3. The spatial distribution maps of pollen percentages for total arboreal pollen (AP) and selected herbaceous taxa (*Artemisia*, Amaranthaceae, Cyperaceae, Poaceae) in the dataset of lake surface-sediment samples (*n*=476) on the Tibetan Plateau.

Group 3 (steppe, *n*=53): *Artemesia* (mean 28.9%, maximum 59.0%) is the most dominant component compared to meadow sites (Fig. 2–4). In addition, as a common taxon, Poaceae (mean 10.3%, maximum 31.4%), as well as the A/C (*Artemisia*/ Amaranthaceae) ratio (range 0.25–12.14, median 3.45), reach their highest values of the different vegetation types.

Group 4 (desert, *n*=10): These sites are characterized by the highest percentages of

删除了: Fig

删除了: Fig

删除了: (=Chenopodiaceae)

Amaranthaceae (mean 26.7%, maximum 52.4%), with higher *Artemisia* abundance (mean 27.4%, maximum 40.2%, Fig. 2–4), and the lowest Poaceae (mean 3.1%, maximum 6.6%), Cyperaceae (mean 11.4%, maximum 21.1%) percentages, and A/C ratio (range 0.55–2.08, median 0.83).

Although AP pollen is detected at most meadow and steppe sites, and occasionally in desert regions, its abundance is markedly lower than that in the forest sites (Table 1, Figs. 2–4). Since trees are absent in the alpine meadow, steppe, and desert communities on the TP (Wu, 1995; Wu and Xiao, 1995; Herzschuh et al., 2010), the low AP abundances likely represent wind-transported pollen transported from adjacent low-elevation regions. Despite this influence, the pollen assemblages effectively represent local vegetation composition, as the contribution of distant pollen is minimal overall (Figs. 2–4). Thus, the modern pollen distribution aligns closely with established vegetation types, corroborating findings from previous studies (Shen et al., 2006; Herzschuh et al., 2010; Li et al., 2020). Pollen assemblages of the 476 pollen samples of the dataset from TP are shown in Figure S1.

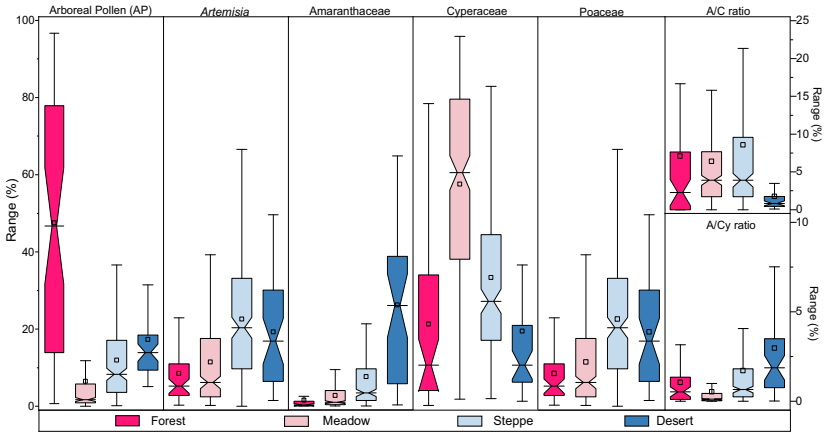

Figure 4. Box plots of the regional percentage distributions of arboreal pollen (AP) and four selected herbaceous pollen types (*Artemisia*, Amaranthaceae, Cyperaceae, Poaceae), plus the ratios of A/C (*Artemisia*/Amaranthaceae (synonym: Chenopodiaceae)) and A/Cy (*Artemisia*/Cyperaceae) from modern lake surface-sediment samples across the Tibetan Plateau.

The initial RDA showed that the VIF values for T_{ann}, Mt_{co}, and Mt_{wa} exceeded 20. Since T_{ann} had the highest VIF, it was removed. After this adjustment, the remaining four variables (NPP, P_{ann}, Mt_{co}, and Mt_{wa}) all had VIF values below 20, and were therefore retained in the final RDA to assess their influence on the modern pollen dataset.

删除了: **Table 2.** Summary statistics of geophysical, climate variables, net primary production (NPP), and pollen percentages of the dataset on the Tibetan Plateau (*n*=476, Min: minimum; Med: median; Max: maximum, SD: standard deviation).
Parameter ... [1]

删除了: (=

删除了: indicat

删除了: variance inflation factor (

删除了:)

删除了: exceeding 20

删除了: the variables

删除了: However, after deleting

删除了: , which

删除了: value,

删除了: , and NPP) had VIF values

删除了: lower than

删除了: . Thus

删除了: they are

删除了: used

删除了: discern

Table 2. Summary statistics of redundancy analysis (RDA) of 476 sites, 35 pollen types, and four climatic variables (P_{ann}: mean annual precipitation, mm; Mt_{co}: mean temperature of the coldest month, °C; Mt_{wa}: mean temperature of the warmest month, °C; T_{ann}: annual mean temperature, °C) and NPP (net primary production) in the pollen dataset from the Tibetan Plateau. VIF: variance inflation factor.

Climatic variables	VIF (without T_{ann})	VIF (with T_{ann})	λ_1/λ_2	Climatic variables as sole predictor	Marginal contribution based on climatic variables	
				Explained variance (%)	Explained variance (%)	P-value
NPP	1.94	2.19	0.21	7.29	0.67	0.006
P_{ann}	3.10	3.43	0.44	13.13	3.92	0.001
Mt_{co}	2.84	80.97	0.09	3.37	2.70	0.001
Mt_{wa}	2.90	41.11	0.15	5.04	1.03	0.001
T_{ann}	—	185.28	—	—	—	—

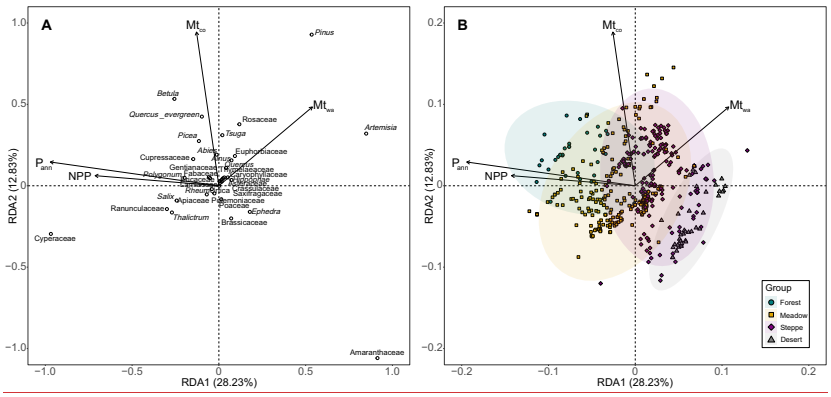

Figure 5. Redundancy analysis (RDA) biplots of the pollen dataset along the first two axes, showing the relationships between (A) 35 selected pollen taxa (circles) and (B) 476 integrated samples (symbols) from different vegetation types and the four variables respectively (arrows): net primary production (NPP, Kg C m^{-2}), mean annual precipitation (P_{ann}, mm), mean temperature of the coldest month (Mt_{co}, °C), and mean temperature of the warmest month (Mt_{wa}, °C).

The RDA results highlight that, as a sole predictor, relative to Mt_{co} and Mt_{wa}, NPP and P_{ann} explain substantial portions of pollen assemblage variation (7.29% and 13.13%, respectively) in the dataset (Table 2). Biplots of the RDA shows that the vectors for both NPP and P_{ann} form smaller angles with the positive direction of axis 1 (capturing 28.23% of total inertia in the dataset) than with axis 2 (12.83%), suggesting moisture availability as the primary determinant along axis 1 (Fig. 5). RDA axis 1, which is highly correlated with NPP and P_{ann}, generally divides the pollen taxa into two groups. One group, comprising Cyperaceae, Ranunculaceae, and *Salix*, indicates wet climatic conditions (located along the positive direction of P_{ann}), while the other group,

删除了: ↵
删除了: **3**
删除了: biplot
删除了: based on
删除了: (
删除了: P_{ann}:
删除了: ; Mt_{co}:
删除了: ,
删除了: ;
删除了: Mt_{wa}:
删除了: ,
删除了: , and NPP: net primary production, Kg C m^{-2}
删除了: P_{ann}
删除了: NPP
删除了: 13.13
删除了: 6.97
删除了: 3
删除了: A
删除了: b
删除了: P_{ann}
删除了: NPP
删除了: 34.7
删除了: compare to
删除了: 15.7

consisting of *Artemisia*, Amaranthaceae, Poaceae, *Ephedra*, and Saxifragaceae represents drought (located along the negative direction of P_{ann}; Fig. 5A). Furthermore, samples collected from alpine desert, steppe, meadow, and forest are located along the gradients of NPP and P_{ann} (Fig. 5B), indicating that they can effectively distinguish different vegetation types as well as pollen assemblages.

5 Potential use of the lake surface-sediment pollen dataset

In the calibration-sets, NPP and P_{ann} are selected as the target variables because of their identified importance in influencing pollen distribution, with NPP further providing insights into alpine vegetation conditions (Table S2). Pollen-based estimates of modern NPP and P_{ann} using both WA-PLS and RF approaches match original measurements well, exhibiting high coefficient of determination (R^2) and low root mean square error of prediction (RMSEP) (Fig. 6). The RF model showed superior predictive performance compared to WA-PLS for both target variables.

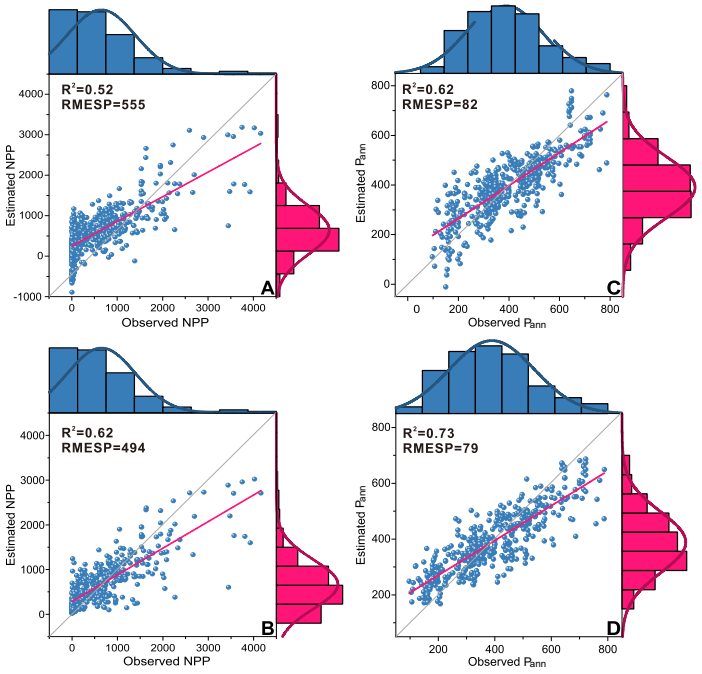

Figure 6. Scatter plots of observed net primary production (NPP) vs. predicted NPP (A, B), observed mean annual precipitation (P_{ann}) vs. predicted P_{ann} (C, D) using weighted-averaging partial least squares regression (WA-PLS: top row) and random forest (RF: bottom row) based on the pollen data (n=476) from lake surface-sediments on the Tibetan Plateau (R^2: coefficient of determination between observed and predicted values; RMSEP: root mean square error of prediction produced by "leave-one-out" cross-validation).

删除了: and NPP
删除了: , as
删除了: The pollen
删除了: P_{ann}
删除了: NPP estimations
删除了: with a
删除了: between observed and predicted variables
删除了: t
删除了: observed net primary production (NPP) vs. predicted NPP …

Reconstructions for NPP (<1000 Kg C m^{-2}) and P_{ann} (ca. 300–600 mm) and are expected to be reliable because their bias is low (Fig. 6). For NPP, RF also shows a notably higher proportion of residuals between -500 and 500 Kg C m^{-2} (84.5%) compared to WA-PLS (74.8%). This advantage persists for the narrower range of -300 to 300 kg C m^{-2} (RF: 63.9% vs. WA-PLS: 50.4%). For P_{ann}, the proportion of residuals between -50 and 50 mm derived from RF (48.1%) is slightly higher than that of WA-PLS (45.6%). Similarly, for the range of -100 to 100 mm, RF (71.8%) outperforms WA-PLS (65.8%). However, both models consistently overestimated NPP and P_{ann} in arid areas with low productivity and underestimated these variables in humid, high-productivity areas, highlighting the necessity of addressing the "edge-effect" (Figs. 6, 7).

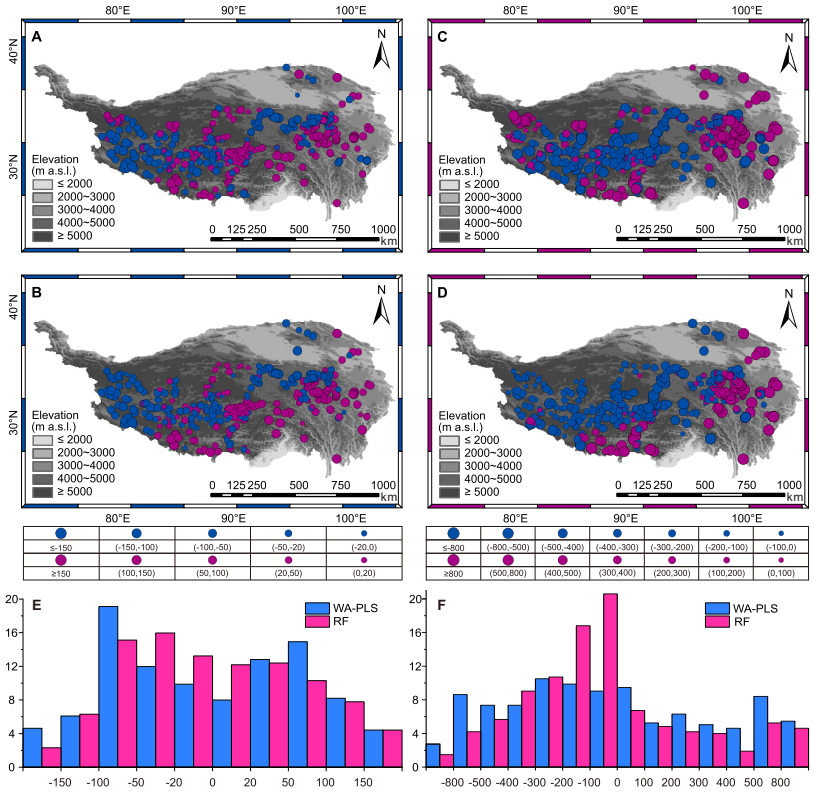

Figure 7. The residuals between observations and pollen-based reconstructions for the lake surface-sediment sites (n=476) on the Tibetan Plateau: net primary production (NPP) by WA-PLS (A) and RF (B), mean annual precipitation (P_{ann}) by weighted-averaging partial least squares regression (WA-PLS) (C) and random forest (RF) (D). The two bar charts in the lower part of the figure show the proportions of modern pollen sites available within different ranges of residuals (observation

删除了: covering
删除了: of
删除了: and NPP lower than 1000 Kg C m^{-2} should be
下移了 [1]: For P_{ann}, the proportion of residuals between -50 and 50 mm derived from RF (48.1%) is slightly higher than that of WA-PLS (45.6%). Similarly, for the range of -100 to 100 mm, RF (71.8%) outperforms WA-PLS (65.8%).
移动了(插入) [1]
删除了: For P_{ann}, the proportion of residuals between -50 and 50 mm derived from RF (48.1%) is slightly higher than that of WA-PLS (45.6%). Similarly, for the range of -100 to 100 mm, RF (71.8%) outperforms WA-PLS (65.8%).
删除了: Pann
删除了: NPP
删除了: Fig
删除了: ↵
... [2]
删除了: (A)
删除了: (B)
删除了:), (C) net primary production (NPP) by WA-PLS and
删除了:) RF.

minus reconstruction) for both NPP (E) and P_{ann} (F).

Most of the poor analogue assemblages come from the TP margin and date back to >12 cal ka BP, possibly due to the higher abundance of arboreal pollen in this specific period and region (Fig. 8). While our combined modern pollen dataset from lake surface-sediments can provide good analogues for fossil pollen assemblages and enhance the performance of palaeoclimate reconstructions on the central TP, caution remains warranted for interpreting pollen assemblages from plateau margins and periods earlier than the Holocene (Fig. 8).

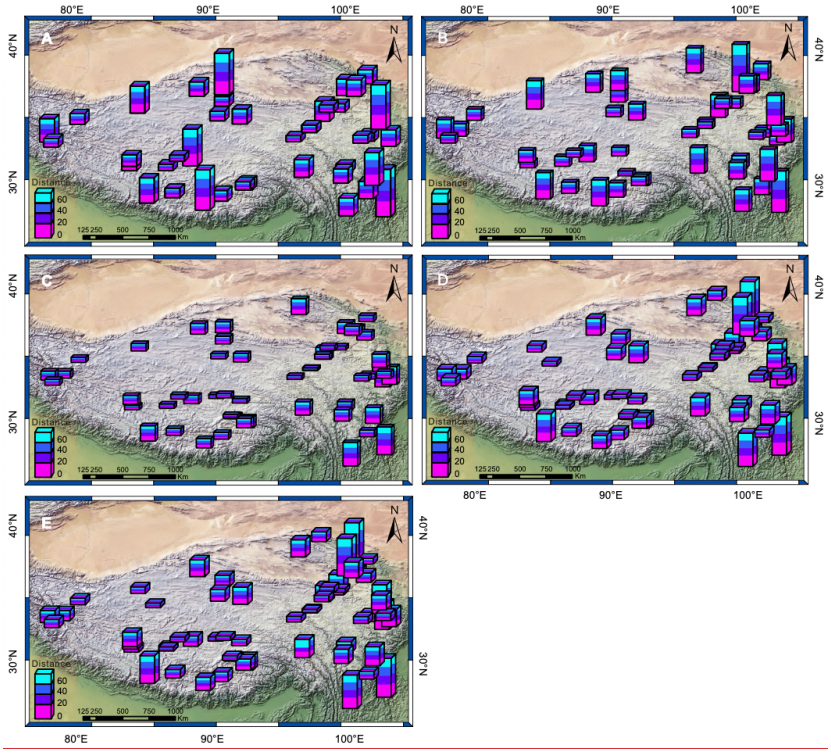

Figure 8. Spatial distribution of analogue quality for six key time slices on the Tibetan Plateau: (A) 15–12 cal ka BP; (B) 12–9 cal ka BP; (C) 9–6 cal ka BP; (D) 6–3 cal ka BP; (E) 3–0 cal ka BP.

6 Data availability

The modern pollen dataset from lake surface-sediment samples (n=90) comprising pollen percentages, site locations, net primary production, and climatic data for each lake is accessible from the National Tibetan Plateau / Third Pole Environment Data Center (TPDC; Tian et al., 2025; https://doi.org/10.11888/Paleoenv.tpdc.302470).

删除了: P_{ann}
删除了: NPP
删除了: which is
删除了: related
删除了: ,

7 Summary

We established a comprehensive modern pollen dataset extracted from lake surface-sediments in forest, meadow, steppe, and desert vegetation types on the TP by combining new modern pollen data with previous datasets. Numerical analyses reveal that P_{ann} is the most important climatic determinant influencing pollen distribution. Our dataset has good predictive power for past NPP and P_{ann} reconstructions. The random forest algorithm is found to be a reliable approach for pollen-based reconstructions of past environments.

The pollen data from our sampled lakes help to fill the geographical gap left by previously published modern pollen datasets, thereby improving the spatial distribution of sampling sites covering the TP. Our dataset is a key component for providing quantitative estimates of past vegetation or climate, and can also be integrated with other pollen datasets in the future to improve the reliability of past ecosystem and climate reconstructions on the TP. Moreover, the current spatial coverage of lakes across the TP is still not fully even, highlighting the need for additional sampling to achieve a more representative dataset in future work.

删除了: Tibetan Plateau

删除了: Tibetan Plateau.

删除了: Tibetan Plateau.

Author contributions. FT and XC designed the pollen dataset, compiled the standardization for the dataset, performed numerical analyses, and organized the manuscript. FT, WC, XC collected the samples, WC performed pollen extraction and identification, XL and ZL prepared the figures and tables. All authors discussed the results and contributed to the final paper.

Competing interests. The corresponding author declares that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors would like to express their gratitude to the palynologists Ulrike Herzschuh (Alfred Wegener Institute Helmholtz Center for Polar and Marine Research), Kai Li (College of Life Sciences, Zhejiang Normal University), Qingfeng Ma (Institute of Tibetan Plateau Research, Chinese Academy of Sciences) who contributed to the dataset. We thank Zhitong Chen (Institute of Tibetan Plateau Research, Chinese Academy of Sciences), and students Meijiao Chen, Yunqing Li, and Anjing Jian for their help with sample collections in the field work.

Financial support. This research was supported by the National Natural Science

Foundation of China (Grant No. 42471179, 42071107).

References

Birks, H. J. B. (Eds.): Quantitative palaeoenvironmental reconstructions. Statistical modelling of Quaternary science data, Vol. 5, Technical guide (ed. by D. Maddy and J.S. Brew), Quaternary Research Association, Cambridge, UK, 271 pp., https://w2.uib.no/filearchive/95birks_qpr_in_maddybrew_1.pdf, 1995.

Birks, H. J. B.: Numerical tools in palaeolimnology–Progress, potentialities, and problems, J. Paleolimnol., 20, 307–332, https://doi.org/10.1023/A:1008038808690, 1998.

Birks, H. J. B., Line, J. M., Juggins, S., Stevenson, A. C., and ter Braak, C. J. F.: Diatoms and pH reconstruction, Philos. T. R. Soc. B, 327, 263–278, https://doi. org/10.1098/rstb.1990.0062, 1990.

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

ter Braak, C. J. F., and Prentice, I. C.: A theory of gradient analysis, Adv. Ecol. Res., 18, 271–317, https://doi.org/10.1016/S0065-2504(03)34003-6, 1988.

ter Braak, C. J. F., and Verdonschot, P. F. M.: Canonical correspondence analysis and related multivariate methods in aquatic ecology, Aquat. Sci., 57, 255–289, https://doi.org/10.1007/BF00877430, 1995.

Breiman, L.: Random Forests, Mach. Learn, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

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., Ni, J., Herzschuh, U., Wang, Y., and Zhao, Y.: A late Quaternary pollen dataset from eastern continental Asia for vegetation and climate reconstructions: Set up and evaluation, Rev. Palaeobot. Palynol., 194, 21–37, https://doi.org/10.1016/j.revpalbo.2013.02.003, 2013.

Cao, X., Tian, F., Andreev, A., Anderson, P. M., Lozhkin, A. V., Bezrukova, E., Ni, J., Rudaya, N., Stobbe, A., Wieczorek, M., and Herzschuh, U.: A taxonomically harmonized and temporally standardized fossil pollen dataset from Siberia covering the last 40 kyr, Earth Syst. Sci. Data, 12, 119–135, https://doi.org/10.5194/essd-12-119-2020, 2020.

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.

Chen, F., Dong, G., Zhang, D., Liu, X., Jia, X., An, C., Ma, M., Xie, Y., Barton, L., Ren, X., Zhao, Z., Wu, X., and Jones, M. K.: Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P., Science, 347, 248–250, https://doi.org/10.1126/science.aaa7573, 2015.

Chen, F., Zhang, J., Liu, J., Cao, X., Hou, J., Zhu, L., Xu, X., Liu, X., Wang, M., Wu, D., Huang, L., Zeng, T., Zhang, S., Huang, W., Zhang, X., and Yang, K.: Climate change, vegetation history, and landscape responses on the Tibetan Plateau during the Holocene: a comprehensive review, Quat. Sci. Rev., 243, 106444, https://doi.org/10.1016/j.quascirev.2020.106444 243, 106444, 2020.

Fægri, K., and Iversen, J. (Eds.): Text book of pollen analysis (4th Edition), John Wiley and Sons

删除了: , 12, 119–135

Press, Chichester, UK, 328 pp., https://doi.org/10.1002/jqs.3390050310, 1989.
Fang, J., Piao, S., Tang, Z., Peng, C., and Ji, W.: Interannual variability in net primary production and precipitation, Science, 293, 1723. https://doi.org/10.1126/ science.293.5536.1723a, 2001.
Gonsamo, A., Chen, J., Price, D. T., Kurz, W. A., Liu, J., Boisvenue, C., Hember, R. A., Wu, C., and Chang, K.: Improved assessment of gross and net primary productivity of Canada's landmass, J. Geophys. Res. Biogeosci., 118, 1546–1560, https://doi.org/10.1002/2013JG002388, 2013.
Grimm, E. C.: CONISS: A FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares, Comput. Geosci., 13, 13–35, https://doi.org/10.1016/0098-3004(87)90022-7, 1987.
Grimm, E. C.: Tilia and Tilia-Graph Software. Springfield, IL: Illinois State Museum [code], https://www.neotomadb.org/apps/tilia, 1991.
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.
Herzschuh, U.: Reliability of pollen ratios for environmental reconstructions on the Tibetan Plateau, J. Biogeogr., 34, 1265–1273, https://doi.org/10.1111/j.1365-2699.2006.01680.x, 2007.
Herzschuh, U., Birks, H. J. B., Mischke, S., Zhang, C., and Böhner, J.: A modern pollen-climate calibration set based on lake sediments from the Tibetan Plateau and its application to a Late Quaternary pollen record from the Qilian Mountains, J. Biogeogr., 37, 752–766, https://doi.org/10.1111/j.1365-2699.2009.02245.x, 2010.
Hill, M. O., and Gauch, H. G.: Detrended correspondence analysis: an improved ordination technique, Vegetatio, 42, 41–58, https://doi.org/10.1007/BF00048870, 1980.
Ji, Y., Zhou, G., Luo, T., Dan, Y., Zhou, L., and Lu, X.: Variation of net primary productivity and its drivers in China's forests during 2000–2018, For. Ecosyst., 7, 15, https://doi.org/10.1186/s40663-020-00229-0, 2020.
Jin, Y., Zhou, K., Gao, J., Mu, S., and Zhang, X.: Identifying the priority conservation areas for key national protected terrestrial vertebrate species based on a random forest model in China, Aeta Ecol. Sin., 36, 7702–7712, 2016. (in Chinese)
Juggins, S.: rioja: Analysis of Quaternary Science Data. version 0.7–3. Juggins, S. [code], https://doi.org/10.32614/CRAN.package.rioja, 2012.
Juggins, S.: Quantitative reconstructions in palaeolimnology: new paradigm or sick science? Quat. Sci. Rev., 64, 20–32, https://doi.org/10.1016/j.quascirev.2012.12.014, 2013.
Juggins, S.: rioja: Analysis of quaternary science data, version 0.9–15.1. Juggins, S. [code], https://doi.org/10.32614/CRAN.package.rioja, 2018.
Juggins, S., and Birks, H. J. B.: Quantitative Environmental Reconstructions from Biological Data. In: Birks, H. J. B., Lotter, A. F., Juggins, S., Smol, J. P. (Eds.), Tracking Environmental Change Using Lake Sediments, Vol. 5: Data Handling and Numerical Techniques, Springer, Dordrecht, Netherlands, 64 pp., https://doi.org/10.1007/978-94-007-2745-8_14, 2012.
Li, C., and Li, Y.: Study of modern pollen and stomata from surficial lacustrine sediments from the eastern edge of Tibetan Plateau, China, Rev. Palaeobot. Palynol., 221, 184–191, https://doi.org/10.1016/j.revpalbo.2015.07.006, 2015.
Li, J., Xie, G., Yang, J., Ferguson, D. F., Liu, X., Liu, H., and Wang, Y. F.: Asian Summer Monsoon changes the pollen flow on the Tibetan Plateau, Earth Sci. Rev., 202, 103114, https://doi.org/10.1016/j.earscirev.2020.103114, 2020.

删除了: Forest Ecosystems 7:

删除了: .

删除了: (23),

Li, X.: Using "random forest" for classification and regression, Chinese J. Appl. Entomol. 50, 1190–1197, 2013. (in Chinese)

Liaw, A.: Random Forest: Breiman and Cutler's Random Forests for Classification and Regression, version 4.6–14, available at: https://cran.r-project.org/web/packages/randomForest/index.html, 2018.

Liu, L., Wang, N., Zhang, Y., Yu, X., and Cao, X.: Performance of vegetation cover reconstructions using lake and soil pollen samples from the Tibetan Plateau. Veg. Hist. Archaeobot., 32, 157–169, https://doi.org/10.1007/s00334-022-00891-0, 2023.

Ma, Q., Zhu, L., Ju, J., Wang, J., Wang, Y., Huang, L., and Haberzettl, T.: A modern pollen dataset from lake surface sediments on the central and western Tibetan Plateau, Earth Syst. Sci. Data, 16, 311–320, https://doi.org/10.5194/essd-16-311-2024, 2024.

Ma, Q., Zhu, L., Wang, J., Ju, J., Lü, X., Wang, Y., Guo, Y., Yang, R., Kasper, T., Haberzettl, T., and Tang, L.: Artemisia/Chenopodiaceae ratio from surface lake sediments on the central and western Tibetan Plateau and its application, Palaeogeogr. Palaeocl. Palaeoecol., 479, 138–145, https://doi.org/10.1016/j.palaeo.2017.05.002, 2017.

Maher, L. J.: Statistics for microfossil concentration measurements employing samples spiked with marker grains, Rev. Palaeobot. Palynol., 32 (2–3), 153–191, https://doi.org/10.1016/0034-6667(81)90002-6, 1981.

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., and Running, S. W.: Climate-driven increases in global terrestrial net primary production from 1982 to 1999, Science, 300, 1560–1563, https://doi.org/10.1126/science.1082750, 2003.

Ni, J.: Carbon storage in Chinese terrestrial ecosystems: approaching a more accurate estimate, Clim. Change, 119, 905–917, https://doi.org/10.1007/s10584-013-0767-7, 2013.

Nychka, D., Furrer, R., Paige, J., and Sain, S.: fields: Tools for spatial data, version 16.3.1 [code], https://doi.org/10.32614/CRAN.package.fields, 2025.

Oksanen, J, Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H.: vegan: Community Ecology Package, version 2.0–4 [code], https://cran.r-project.org/web/packages/vegan/index.html, 2019.

Pepin, N., Bradley, R. S., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M. Z., Liu, X., Miller, J. R., Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schöner, W., Severskiy, I., Shahgedanova, M., Wang, M. B., Williamson, S. N. and Yang, D.: Elevation-dependent warming in mountain regions of the world, Nature Clim. Change, 5, 424–430, https://doi.org/10.1038/nclimate2563, 2015.

Prentice, I. C.: Multidimensional scaling as a research tool in Quaternary palynology: a review of theory and methods, Rev. Palaeobot. Palynol., 31, 71–104, https://doi.org/10.1016/0034-6667(80)90023-8, 1980.

Qin, F.: Modern pollen assemblages of the surface lake sediments from the steppe and desert zones of the Tibetan Plateau, Sci. China Earth Sci., 64, 425–439, https://doi.org/10.1007/s11430-020-9693-y, 2021.

Qin, F., Zhao, Y., and Cao, X.: Biome reconstruction on the Tibetan Plateau since the Last Glacial Maximum using a machine learning method, Sci. China Earth Sci., 65, 518–535, https://doi.org/10.1007/s11430-021-9867-1, 2022.

R Core Team: R, A language and environment for statistical computing, R Foundation for Statistical

删除了: and

删除了: .

Computing [code], https://www.r-project.org, 2019.
Shen, C., Liu, K., Tang, L., and Overpeck, J. T.: Quantitative relationships between modern pollen rain and climate in the Tibetan Plateau, Rev. Palaeobot. Palynol., 140, 61–77, https://doi.org/10.1016/j.revpalbo.2006.03.001, 2006.
Sun, H. (Eds.): The national physical atlas of China, China Cartographic Publishing House, Beijing, China, 283 pp., ISBN 9787503120398, 1999.
Tang, L., Mao, L., Shu, J., Li, C., Shen, C., and Zhou, Z. (Eds.): An Illustrated Handbook of Quaternary Pollen and Spores in China, Science Press, Beijing, China, 620 pp., ISBN9787030505682, 2016. (in Chinese)
Telford, R. J., and Birks, H. J. B.: Effect of uneven sampling along an environmental gradient on transfer-function performance, J. Paleolimnol., 46, 99–106, https://doi.org/10.1007/s10933-011-9523-z, 2011.
Telford, R. J.: palaeoSig: significance tests for palaeoenvironmental reconstructions, version 1.1–2, Telford, R. J. [code], https://doi.org/10.32614/CRAN.package.palaeoSig, 2013.
Tian, F., Cao, X., Zhang, R., Xu, Q., Ding, W., Liu, X., Pan, B., and Chen, J.: Spatial homogenization of soil-surface pollen assemblages improves the reliability of pollen-climate calibration-set, Sci. China Earth Sci., 63, 1758–1766, https://doi.org/10.1007/s11430-019-9643-0, 2020.
Tian, F., Cao, W., Liu, X., Liu, Z., and Cao, X.: Pollen assemblages of lake surface sediment across the Tibetan Plateau, National Tibetan Plateau / Third Pole Environment Data Center [data set], https://doi.org/10.11888/Paleoenv.tpdc.302470, 2025.
Tian, F., Wang, W., Rudaya, N., Liu, X., and Cao, X.: Wet mid–late Holocene in central Asia supported prehistoric intercontinental cultural communication: Clues from pollen data, Catena, 209, 105852, https://doi.org/10.1016/j.catena.2021.105852, 2022.
Walker, A. P., Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Asao, S., Hickler, T., Parton, W., Ricciuto, D. M., Wang, Y., Wårlind, D., and Norby, R. J.: Predicting long-term carbon sequestration in response to CO_2 enrichment: how and why do current ecosystem models differ? Global Biogeochem. Cycles, 29, 476–495, https://doi.org/10.1002/2014GB004995, 2015.
Wang, B. (Eds.): The Asian Monsoon, Springer, Chichester, UK, 845 pp., https://doi.org/10.1007/3-540-37722-0, 2006.
Wang, F., Qian, N., Zhang, Y., and Yang, H. (Eds.): Pollen Flora of China, 2nd Edition, Science Press, Beijing, China, 461 pp., ISBN7030036352, 1995. (in Chinese)
Wu, K., Li, K., Jia, W., Stoof-Leichsenring, K. R., Herzschuh, U., Ni, J., Liao, M., and Tian, F.: Application of plant DNA metabarcoding of lake sediments for monitoring vegetation compositions on the Tibetan Plateau, Sci. China Earth Sci., 67, 3594–3609, https://doi.org/10.1007/s11430-023-1358-0, 2024.
Wu, Y., and Xiao, J.: A preliminary study on modern pollen rain of Zabuye Salt Lake area, Xizang, Plant Divers., 17, 72–78, https://journal.kib.ac.cn/EN/Y1995/V17/I01/1, 1995. (in Chinese)
Wu, Z. (Eds.): The vegetation of China, Science Press, Beijing, China, 1270 pp., ISBN 7030024222, 1995. (in Chinese)
Yang, G. (Eds.): China lake survey, Science Press, Beijing, China, 671 pp., ISBN 9787030614223, 2019. (in Chinese)
Yue, P., Lu, X., Ye, R., Zhang, C., Yang, S., Zhou, Y., and Peng, M.: Distribution of Stipa purpurea steppe in the Northeastern Qinghai-Xizang Plateau (China), Russ. J. Ecol., 42, 50–56, https://doi.org/10.1134/S1067413611010140, 2011.

Zhang, X.: Vegetation Map of China and Its Geographic Pattern-Illustration of the Vegetation Map of the People's Republic of China (1:1000000), Geology Press, Beijing, China, https://doi.org/10.12282/plantdata.0155, 2007. (in Chinese)

Zhao, M., and Running, S. W.: Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009, Science, 329, 940–943, https://doi.org/10.1126/science.1192666, 2010.

第 11 页: [1] 删除了　　匿名　　2025/8/29 21:22:00

第 14 页: [2] 删除了　　匿名　　2025/8/29 21:22:00