Science Bulletin 67 (2022) 1115–1117

Contents lists available at ScienceDirect

Science Bulletin

journal homepage: www.elsevier.com/locate/scib

Commentary

Long-distance modern analogues bias results of pollen-based precipitation reconstructions

Xianyong Cao [a,*], Jianhui Chen [b], Fang Tian [c], Qinghai Xu [d], Ulrike Herzschuh [e,f,g], Richard Telford [h], Xiaozhong Huang [b], Zhuo Zheng [i,j], Caiming Shen [k], Wenjia Li [a,l]

[a] Alpine Paleoecology and Human Adaptation Group, State Key Laboratory of Tibetan Plateau Earth System, and Resources and Environment, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China
[b] Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China
[c] College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China
[d] School of Geographic Sciences, Hebei Normal University, Shijiazhuang 050024, China
[e] Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Research Unit Potsdam, Potsdam 14473, Germany
[f] Institute of Environmental Science and Geography, University of Potsdam, Potsdam 14476, Germany
[g] Institute of Biochemistry and Biology, University of Potsdam, Potsdam 14476, Germany
[h] Department of Biological Sciences & Bjerknes Centre for Climate Research, University of Bergen, Bergen N-5020, Norway
[i] Guangdong Key Lab of Geodynamics and Geohazards, School of Earth Sciences and Engineering, Sun Yat-sen University, Zhuhai 519082, China
[j] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China
[k] Yunnan Key Laboratory of Plateau Geographical Processes and Environmental Changes, Faculty of Geography, Yunnan Normal University, Kunming 650050, China
[l] University of Chinese Academy of Sciences, Beijing 100049, China

Timing of the precipitation optimum in the Holocene for the semi-arid northern China affects our understanding of the temporal-patterns of the East Asian Summer Monsoon and its connection with precipitation in the monsoon fringe area. Discrepancies about when this occurs (early Holocene or mid-Holocene) exist in paleoclimate records based on various proxies and models. What causes these discrepancies is still a hot topic [1]. Past annual precipitation (P_{ann}) was reconstructed from Holocene pollen spectra extracted from Gonghai Lake by Cheng et al. [2] using a subcontinental scale modern pollen dataset (n = 1865, mostly covering China and Mongolia) and the modern analogue technique (MAT) to reveal the P_{ann} temporal patterns for semi-arid northern China and to assess the reliability of model estimations and the uncertainty of pollen-based reconstructions. Cheng et al.'s [2] P_{ann} reconstruction has an early-Holocene maximum, contrasting with previous P_{ann} reconstructions from the same pollen spectra [3] and from other Holocene fossil pollen spectra spread across northern China [4], which all show a mid-Holocene maximum. After investigating the limitations of MAT and the reconstruction processes undertaken by Cheng et al. [2], we conclude that their P_{ann} reconstruction methodology yielding an early-Holocene maximum is not reliable.

The basic assumption when applying MAT is that similar biological assemblages are deposited under similar environmental conditions, such that the environment of a fossil sample can be estimated by comparison to its modern analogues [5,6]. To produce a reliable reconstruction, MAT requires a reasonable coverage of modern environmental conditions within a restricted (bio-)geographic space, because a calibration-set with a large spatial extent will increase the risk of "multiple analogues", where similar modern biological assemblages reflect different environmental conditions and thus bias the reconstruction [5,6]. The limited taxonomic resolution of pollen identification (generally at genus or even family level) also implies that the risk of "multiple analogues" is higher for a pollen-based environmental reconstruction than for other biological data, particularly in an area with high plant diversity (e.g., eastern Asia) [7]. Therefore, for a pollen-based past climate reconstruction, particularly when using MAT, the spatial extent of the calibration-set ought to be reasonably restricted [7] and/or pollen taxa separated into different regional groups [8] to avoid the risk of "multiple analogues".

The subcontinental scale modern pollen dataset employed by Cheng et al. [2] covers various (bio-)geographic and climate zones characterized by a high plant diversity [9]. In the dataset, various plant species that differ in their climate envelopes are represented by a single pollen taxon, for example, *Pinus*, *Picea*, *Betula* and *Quercus* [10]. Even worse, pollen taxa with similar phenological and climate constraints were combined by Cheng et al. [2] into the same plant functional type (PFT; for instance, *Abies*, *Picea* and *Pinus* were combined into a PFT named by boreal evergreen conifer) for calculating the dissimilarity between fossil and modern data [2]. This approach improves the "analogue quality" superficially, while simultaneously increasing the risk of false matches [5,7,8]. Hence, MAT is not the optimal option for this subcontinental modern pollen dataset with low taxonomic resolution, and the P_{ann} reconstruction showing an early-Holocene maximum completed by Cheng et al. [2] bears a particularly high risk of being based on "multiple analogues".

Unfortunately, Cheng et al. [2] did not share their modern pollen dataset in their supplementary materials, thus we constructed

* Corresponding author.
E-mail address: xcao@itpcas.ac.cn (X. Cao).

https://doi.org/10.1016/j.scib.2022.01.003

a modern pollen dataset [11] (n = 2559, climate data obtained from WorldClim climate data version 2 following Cheng et al. [2]) from the same geographic and climatic ranges (likely containing mostly the same samples) to test the hypothesis that temporal occurrence of the Holocene P_{ann}-maximum in samples from northern China depends on the spatial extent of the modern dataset. Two pollen–P_{ann} calibration-sets were established by MAT based on original pollen percentages of (1) all available sites ("large set") and (2) sites within a 1000-km radius around the Gonghai Lake ("small set"), which were then used to reconstruct P_{ann} for the Gonghai pollen spectra. For comparison, weighted-averaging partial least squares (WA-PLS) were also employed in P_{ann} reconstructions using the two calibration-sets. The better model performance of cross-validation for MAT (Fig. S1 online) is caused by strong spatial autocorrelation [11], hence we argue that the WA-PLS model based on the "small set" is actually better but still has reasonable analogue quality (Fig. S2 online). Detailed information about data processing, pollen-climate relationship investigations, and model assessments together with reconstructions are presented in Cao et al.'s [7,11] papers.

In accordance with Cheng et al.'s [2] results (Fig. 1a), the P_{ann} reconstruction produced by MAT using the "large set" presents an early-Holocene maximum. However, a relatively dry early Holocene and mid-Holocene optimum were estimated by both MAT and WA-PLS using the "small set" and by WA-PLS with the "large set" (Fig. 1b, d and c, respectively), which is consistent with previous climate reconstructions for northern China [3,4]. The greatest difference in reconstructed P_{ann} between the two approaches is obtained for fossil pollen spectra from 10–7 cal ka BP (Fig. 1a), where the abundance of *Betula* and *Picea* reach a maximum and thus strongly impact the reconstructed P_{ann} (Fig. S3 online). The origins of their modern analogues can be used as a direct and effective approach to explain the difference revealed by the two calibration-sets. The modern analogues for these fossil samples within 10–7 cal ka BP selected from the "large set" are located mainly in North-East China (the Changbai Mts.) and on the eastern Tibetan Plateau where modern P_{ann} (mean = 615 mm) is higher than for those selected from the "small set" (from mountainous areas around Gonghai Lake; mean = 518 mm) (Fig. 1e, f). The pollen-climate relationship reveals that *Picea* and *Betula* have higher P_{ann}-optima on the eastern Tibetan Plateau than in North-East China [10]. In addition, *Picea* includes species with different P_{ann} ranges in the three geographic regions. For instance, in northern China (where Gonghai Lake is located), *Picea wilsonii* is common (P_{ann} range 132–1358 mm, mean = 602 mm); while *P. jezoensis* is common in the Changbai Mountains (561–1073 mm; mean = 789 mm); and on the eastern Tibetan Plateau, *P. brachytyla* (895–1435 mm; mean = 895 mm) and *P. likiangensis* (320–1431 mm; mean = 699 mm) are typically found [12] (Fig. S4 online). Therefore, we conclude that the reconstructed early-Holocene P_{ann} maximum is caused by these long-distance analogues that represent a quite different vegetation to those that likely occurred around Gonghai Lake in the past. For fossil pollen spectra between 6 and 0 cal ka BP, although there are many long-distance analogues from the "large set", the analogues selected overlap in both sets ensuring the similarity between the two reconstructions since 6 cal ka BP (Fig. S5 online). Our investigation confirms the P_{ann} record reconstructed by Cheng et al. [2] is likely affected strongly by the well-known "multiple analogues" problem due to employment of a too extensive spatial calibration-set with quite low taxonomic resolution; and the difference in results reconstructed by Cheng et al. [2] and Chen et al. [3] is caused by different selections of calibration-set rather than approaches.

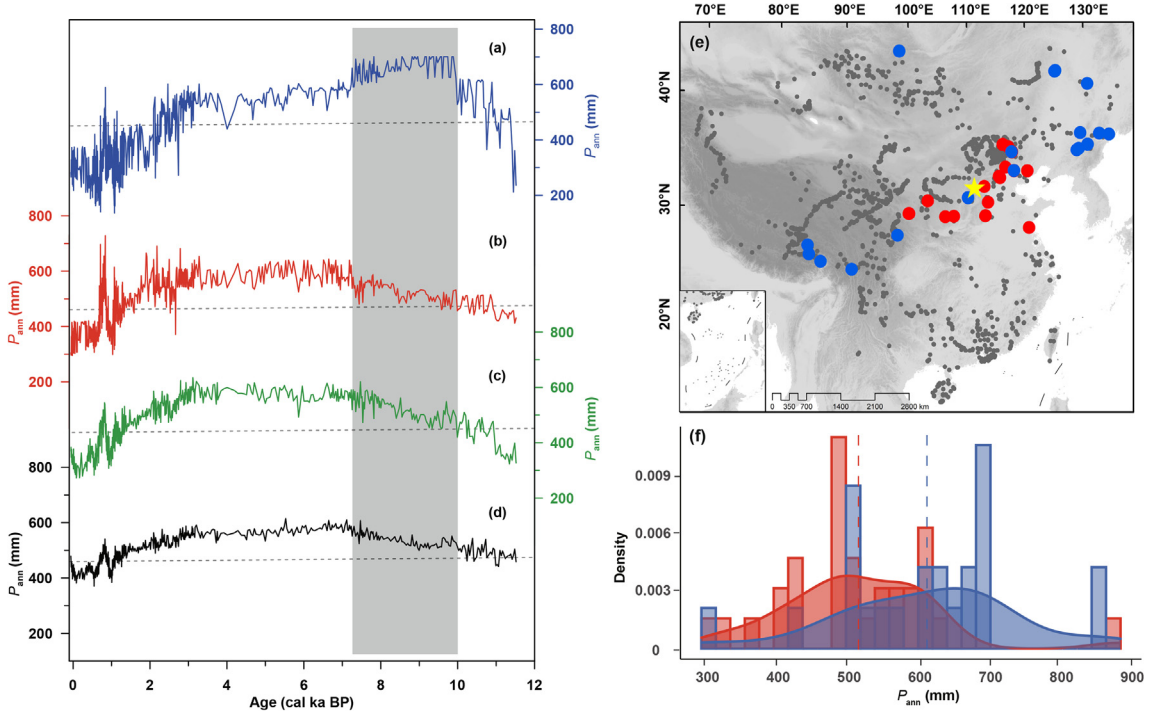

Fig. 1. Holocene annual precipitation (P_{ann}) reconstructions based on original pollen percentages of all available sites ("large set"; (a) blue curve based on MAT, (c) green curve based on WA-PLS) and sites within a 1000-km radius around the Gonghai Lake ("small set"; (b) red curve based on MAT, (d) black curve based on WA-PLS); the grey horizontal dashed lines indicate the modern observed P_{ann} for Gonghai area. (e) Locations of modern analogues for the fossil pollen spectra between 10 and 7 cal ka BP selected for the "large set" (blue dots) and the "small set" (red dots) determined by MAT with the squared-chord distance, together with the location of Gonghai Lake (yellow star) and all available pollen sites (grey dots). (f) P_{ann} summary for the long-distance analogues selected from the "large set" (blue) and the nearby analogues from the "small set" (red); the vertical dashed lines are the means for these two groups' analogues.

We therefore propose that P_{ann} reconstructions using WA-PLS based on a spatially restricted "small set" are more reliable than reconstructions with a "large set" when using MAT, because pollen taxa – in particular woody taxa – contribute strongly to the Gonghai Lake P_{ann} reconstruction and show rather stable distribution ranges during the Holocene (i.e., migrations of most woody taxa are spatially only limited during the Holocene) [13]. Additionally, the general temporal pattern of P_{ann} reconstruction based on the "small set" is consistent with the modern observed P_{ann} in the Gonghai region, environmental reconstructions by multiple proxies for Gonghai Lake [3,14] and the semi-arid northern China [4], and the reconstructions based on WA-PLS using both the "large set" and "small set" (Fig. 1c, d) (results comparison with other methods is particularly necessary for MAT using large training-sets [5,6]).

We conclude that trends and quantities of pollen-based P_{ann} reconstructions from northern China using a subcontinental scale calibration-set within a MAT approach are unreliable, and the problem exacerbated when the taxonomic resolution is artificially lowered such as with the PFT approach employed by Cheng et al [2]. We consider that the high and stable P_{ann} reconstructions between ca. 10 and 3 cal ka BP as inferred by Cheng et al. [2] results from using distant analogues from the Changbai Mountains and these analogues should be utilized repeatedly.

Our finding is in agreement with previous conclusions that pollen-based environmental reconstruction is heavily dependent on the chosen numerical methods, spatial extent, and taxonomic resolution of the modern pollen dataset, as well as on the suitability of the modern pollen assemblages for the fossil assemblages [7,15]. Hence, prior to reconstruction, a series of statistical analyses need to be performed to select the most suitable target variable, the most robust method, and the most suitable modern dataset; and finally, the reconstruction should be assessed.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (41988101 and 41877459), and the Pioneer Hundred Talents Program of Chinese Academy of Sciences. Drs. Haichao Xie and Jie Liang (Institute of Tibetan Plateau Research, Chinese Academy of Sciences) provided help with the figure preparation; Cathy Jenks (University of Bergen) provided help with language editing. Fossil and modern pollen data employed in this study are available as Supplementary materials to Herzschuh et al. [4].

Appendix A. Supplementary materials

Supplementary materials to this commentary can be found online at https://doi.org/10.1016/j.scib.2022.01.003.

References

[1] Liu X, Liu J, Chen S, et al. New insights on Chinese cave $\delta^{18}O$ records and their paleoclimatic significance. Earth-Sci Rev 2020;207:103216.
[2] Cheng J, Wu H, Liu Z, et al. Vegetation feedback causes delayed ecosystem response to East Asian summer monsoon rainfall during the Holocene. Nat Commun 2021;12:1843.
[3] Chen F, Xu Q, Chen J, et al. East Asian summer monsoon precipitation variability since the last deglaciation. Sci Rep 2015;5:11186.
[4] Herzschuh U, Cao X, Laepple T, et al. Position and orientation of the westerly jet determined Holocene rainfall patterns in China. Nat Commun 2019;10:2376.
[5] Juggins S, Birks HJB. Quantitative environmental reconstructions from biological data. In: Birks HJB, Lotter AF, Juggins S, et al., editors. Tracking environmental change using lake sediments (Volume 5): data handling and numerical techniques. Dordrecht: Springer; 2012. p. 431–94.
[6] Birks HJB, Heiri O, Seppö H, et al. Strengths and weaknesses of quantitative climate reconstructions based on late-Quaternary biological proxies. Open Ecol J 2010;3:68–110.
[7] Cao X, Tian F, Telford RJ, et al. Impacts of the spatial extent of pollen-climate calibration-set on the absolute values, range and trends of reconstructed Holocene precipitation. Quat Sci Rev 2017;178:37–53.
[8] Williams JW, Shuman B. Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. Quat Sci Rev 2008;27:669–87.
[9] Zheng Z, Wei J, Huang K, et al. East Asian pollen database: modern pollen distribution and its quantitative relationship with vegetation and climate. J Biogeogr 2014;41:1819–32.
[10] Tian F, Cao X, Dallmeyer A, et al. Pollen-climate relationships in time (9 ka, 6 ka, 0 ka) and space (upland vs. lowland) in eastern continental Asia. Quat Sci Rev 2017;156:1–11.
[11] Cao X, Herzschuh U, Telford RJ, et al. A modern pollen-climate dataset from China and Mongolia: assessing its potential for climate reconstruction. Rev Palaeobot Palynol 2014;211:87–96.
[12] Fang J, Wang Z, Tang Z. Atlas of woody plants in China: distribution and climate (Volume I). Beijing: Higher Education Press, and Berlin: Springer; 2011.
[13] Cao X, Herzschuh U, Ni J, et al. Spatial and temporal distributions of major tree taxa in eastern continental Asia during the last 22,000 years. Holocene 2015;25:79–91.
[14] Chen S, Liu J, Xie C, et al. Evolution of integrated lake status since the last deglaciation: a high-resolution sedimentary record from Lake Gonghai, Shanxi, China. Palaeogeogr Palaeoclimatol Palaeocol 2018;496:175–82.
[15] Telford RJ, Birks HJB. A novel method for assessing the statistical significance of quantitative reconstructions inferred from biotic assemblages. Quat Sci Rev 2011;30:1272–8.

Xianyong Cao was awarded the Ph.D. degree of paleoecology in 2015 from University of Potsdam, Germany. He has been a professor at Institute of Tibetan Plateau Research, Chinese Academy of Sciences since 2018. His research focuses on pollen-based past vegetation and climate quantitative reconstruction (late Quaternary) at multi-spatial scales.