

Response to the reviewer #2

We have responded to all the comments and questions, answering them one by one. Our responses are in blue script placed after the reviewers' comments.

Reviewer #2

I am grateful for the opportunity to review the article entitled "Examining bias in pollen-based quantitative climate reconstructions induced by human impact on vegetation" by Wei Ding, Qinghai Xu, Pavel Tarasov, for *Climate of the Past*.

I think this paper is well written, very clear and concise, and sufficiently interdisciplinary in content to be published in *Climate of the Past*. This study focuses on two modern pollen data calibration sets, one considered as "natural", one with high human impact. These 2 modern pollen datasets are tested to reconstruct the annual precipitation changes during the late glacial and the Holocene in north-central China, at the margin of the East Asian Summer; they used the classical Weighted Averaging Partial Least Squares (WA-PLS) approach to reconstruct quantitatively the precipitation.

This study is very interesting given that the problem of human impact on pollen data is a key problem in climate reconstructions. I think that the paper of Ding et al. presents interesting findings in terms of results; in particular they support the consensus that climate reconstruction based on pollen data must be taken with caution for recent periods (Bronze Age to 0 ka).

So, I recommend the publication with minor changes, which are listed below.

Main points

- I think that test two different datasets -, one considered as "natural", one with high human impact - is really interesting. But in fact, transfer functions to reconstruct past climate are based on the principle that vegetation and climate are in equilibrium-which is not always true, I know-. We keep only pollen samples which have been collected in « natural » ecosystems, we exclude samples which have been collected in areas with high human impact. One another test interesting to do is to test the reconstruction with an unique dataset (natural and non natural) to test if the results will be different or not because we will never use a dataset with human impact.

Indeed, scientists normally avoid collecting reference samples from human-induced vegetation for establishing their pollen-climate calibration set. The difference between reconstructed results using all samples and 'good' samples is often tiny as only a few samples with obvious human impact are excluded, and the model mechanism is not clear, although the statistical performance is better for using 'good' samples (Xu et al., 2010). Based on a large size data set using human-impacted pollen samples compiled from our previous works, we have a chance to make a control-group to test the performance of 'nature' samples. The reconstructed results (value and statistical performance) using all samples (natural and non-natural) occur between the results using the natural set and the human-induced set. For clearly

showing the model-inferred optimum altered by human impact and the largest potential bias, we chose to show the results using the natural set and human-induced set though the latter one is not commonly used in a reconstruction work.

- what is the definition of « natural vegetation » given the changes made by human societies on their environment, particularly during the late Holocene? How do you define natural vegetation? is it potential vegetation?

For most parts in the temperate eastern China, the human influences are more intense than before, hence, it is impossible to establish a surface pollen data set with samples from completely natural vegetation, no matter since when human societies had started to change the vegetation. Here, 'nature vegetation' was more defined in practice instead of in theory. We classify the samples to nature group in field work according to one or more following criteria: collecting samples (1) in natural reserve, (2) in areas where vegetation composition is coincident with its potential scenario, (3) from vegetation community where is hard for human access and without obvious synanthropic plants, (4) away from settlements, roads, paths, farmlands, pasture lands, and other human activities as far as possible.

- your study takes only into account the bias linked to modern pollen datasets; what about the potential biases on fossil data linked to past human impact (fires ...); could you more discuss this point?

Examining the biases introduced from past human impact is one task of our future works. In this study, we have shown the potential biases during the late Holocene, especially since the Bronze Age. Archeological and historical documents, and even cereal-type pollen evidence are indeed existed, but we have difficulty in quantitatively separating human impact from natural change. Based on the results with our contrasting data sets, it seems that we still can extract plausible climate information (at least for the tendency) from fossil pollen data, nevertheless, the pollen-based climate reconstructions in temperate eastern China should be conducted with more caution and a possible bias range should be kept in mind for further discussions.

- I have not seen a description of the fossil record used. Even if it has been already published, it's an important point. Could you add a simplified pollen diagram for the fossil pollen sequence? Could you give details on pollen assemblages? Or describe it very briefly? Time period covered?

We have added the fossil pollen diagram and brief introduction in both text and figure outline.

*In section 3.3, we add a paragraph: "Vegetation succession in the area around Lake Gonghai has experienced five major stages during the last 14.7 ka (Fig. 2). Open forests and upland meadows dominated during the last deglaciation (14.7–11.1 ka) and abrupt strengthening of *Artemisia*-dominated mountain steppe association occurred during 13.1–12.0 ka BP. In the early Holocene (11.1–9.6 ka), *Betula*, *Carpinus*, *Ostryopsis*, and *Ulmus* as pioneer tree species spreaded in the landscape. Later (9.6–7.3 ka) mixed forest dominated by *Picea* and *Betula* started to play a greater role. Temperate deciduous trees (e.g. *Quercus*) widely expanded and the mixed broadleaf-conifer forest grew around the lake during the Holocene climatic optimum (7.3–5.0 ka BP). The break up of the climax community started from 5.0 ka BP with the increasing of *Pinus* and the decreasing of *Quercus* and *Betula* percentages. Major increase*

in herbaceous pollen percentages occurred since 2.8 ka BP, especially during the last 1.6 ka, when *Humulus* (including *Urtica*), *Fagopyrum*, and cereal Poaceae pollen types related to human activities became most prominent in the diagram (Fig. 2; Xu et al., 2016). Vegetation dynamic inferred from GH09B pollen sequence is a valuable source of environmental information for the EASM margin (Chen et al., 2015). In the current study, we selected this well-dated and high-resolution fossil record to reconstruct past climate, using both the N-set and H-set of modern pollen as calibration sets.”

The new figure was marked as Figure 2 (see below).

- Did you calculate error bars? I have not seen them in the figure. Error bars are needed before publication to discuss more in depth the possible bias with the H set. And to see if the differences between the 2 climate reconstructions are significant or not.

We have added the error bars (standard error of prediction) for the reconstructed P_{ann} values in Figure 6 (as Figure 4 in the previous version of the manuscript, see below).

Minor points

1. Title: could you add “in China” in the title?

Added.

2. The abstract is well written and informative. Maybe you can precise in the abstract that the bias in the pollen-based reconstruction is important only during recent periods (not entire lateglacial and Holocene)

We added this conclusive information in the end of the abstract.

“The extent of human-induced bias may be rather small for the entire lateglacial and early Holocene interval when we use a reference set called ‘natural’. Nevertheless, this potential bias should be kept in mind when conducting quantitative reconstructions, especially for recent two or three millennia.”

3. Introduction

- p.1, line 26: the ref Von Post, (1916) is more appropriate

Changed.

- p 1, line 30: I don’t agree with the sentence: Palaeoclimatology relies on modern pollen-climate relationship studies; Palaeoclimatology is based on pollen but also on various proxies (speleothems, lakes, tree-rings...) not only pollen; please correct.

Corrected. We have rewritten this sentence as “Pollen-based palaeoclimate reconstructions rely on modern pollen-climate relationship studies”.

- p.1, line 32-33: the ref Bartlein et al., 2011 is missing

Added.

- p.1, line 36: what is” the principle of uniformitarianism”?

We have rewritten this phrase as: “A methodological assumption that the ecological response of species does not change during the Quaternary based on Lyell's uniformitarianism (Scott, 1963)”.

- p.2, line 5: “in such regions”: which ones?

This phrase “in such regions” here means “In China and other regions with long-term human occupation” hereinbefore, and we have rewritten the sentence as: It is most likely, that modern pollen-climate relationships in such regions with long-lasting human impact are different from what they were in the past.

4. Material and methods

- p.3, line 28: how many pollen data do you use from each region and ecosystem types?

We have added this information in the text.

“...including 43 spectra from Anyang area, central China Plain (Wang et al., 2009), 12 from eastern Hexi Corridor (Ma et al., 2009), 78 from warm temperate hilly areas (Ding et al., 2011), 88 from Hebei Plain and adjacent mountains area (Pang et al., 2011), 13 from south-east China (Yang et al., 2012), and 105 spectra from north-east China (Li et al., 2012; Li et al., 2015).”

- p.3, line 37: how do you define the “natural vegetation communities? Do you have pollen traps and vegetation surveys?

We defined the vegetation as “natural” in some sites based on trap samples but mostly relied on vegetation surveys (e.g. Li et al., 2011). As mentioned above in our answer to main point 2, “natural” here was used to refer these vegetation communities which keep their potential characteristics in structure and composition, and without obvious human disturbance.

- p.4, line 6: is the pollen sum enough high after exclusion of anthropic taxa, for example in cultivated lands? line 7: it will be informative to precise how many pollen spectra you have per ecosystem (both dataset).

The minimum number of pollen grains to count is 400 in our works, and 500~600 pollen grains were counted for most samples. The mean percentages of cultivated pollen taxa for different type samples are 5~25%, and only up to 60% in some farmland samples. After excluding the human-indicator taxa, most samples still have enough grains (300~500) for further calculations. However, samples with very high percentage of cereal-type pollen were not used in our study.

We have added the samples-ecosystem information in the text: “The N-set includes 11 samples from vegetation region I, 61 from region II, 300 from region III, 351 from region VI, 25 from region IVAi, 51 from region VIIBi, and 7 from region VIIIAi, while 14 samples from region I, 11 from region II, 433 from region III, 292 from region VI, and 41 from region IVAi appear in the H-set.”

- p.4, lines 11-12: do you also have calculated GDD5 and the moisture index (prentice et al...)?

Yes, we had tested the GDD5 and a moisture index (the ratio of evapotranspiration to potential evapotranspiration, AET/PET, as effective moisture, water availability or a index in different literatures) which both are important bioclimatic variables for vegetation distribution. However, the GDD5 is not a determinant (control) variable in the study area and its explanatory power is not better than thermal variables used in the study. For effect moisture, MODIS-based AET/PET data from (Zhao and Running, 2010) was used. It is highly correlated to P_{ann} because our study area mainly including sub-humid and semi-arid regions where P_{ann} is a determinant variable. Therefore, we did not use these two variables in our study.

- p.4, line 14: how do you correct the values for the precipitation parameter?

We did not correct the precipitation values with any calibration. The precipitation lapse rate is still not available in our study area. There are some results from sites studies in western China (e.g. the precipitation value increases with the elevation in Tianshan area and the multi-year average rate are 6~11 mm per 100m for different slopes), however, we are not sure it can be used in monsoon region or not. Our pollen-climate data set is the first time to use new available and high-resolution observation data from 1208 meteorological stations, which are well-distributed in the study area and cover the elevation range of 0~4000 m. The interpolation of precipitation values for pollen sites is credible.

- p.4, lines 18: more details on the calculation of HII are needed: HII is based only on modern human impact, it doesn't take into account the human impact during recent periods (from 2800 to 0).

We have rewritten the paragraph about HII in section 3.2:

Sanderson et al. (2002) developed a human influence index (HII) dataset for mapping the areas with and without human footprint. The HII dataset quantifies human influence on terrestrial ecosystems based on four proxies (nine datasets), including human population pressure (population density), land transformation (land use/cover, roads and railways, built-up centers, settlements), accessibility (roads and railways, coastlines, navigable rivers) and electrical power infrastructure (night-time lights). Each of the nine datasets assigns a score from 0 to 10 according to a rating (or alternative in a single score and 0) to assess human influence on 1-km² land surface. Sum scores from nine datasets were standardized as HII values which range from 0 to 64 (WCS/CIESIN, 2005), with higher value indicating higher degree of human impact (Fig. 1b). HII reflects only modern human impact and doesn't consider the human impact during the past. The HII has been recently employed to assess human influence on pollen assemblages in China (Li et al., 2014; Li et al., 2015). In this study, HII values at 1-km² grids are assigned to modern pollen reference sites using ArcGIS.

- p. 4 part 3.3 see main point 4

We have added the pollen diagram and a paragraph in the text (See above).

- p.5, line 3: which statistical techniques have been tested? Could you write 1 sentence on the concept of the WAPLS?

We have rewritten the sentences as: "The WA-PLS approach combines the virtues of WA method to

model ecological optima of species and PLS method to select linear components from biological assemblages (ter Braak and Juggins, 1993). It has been tested, along with WA, MAT, and pollen response surface method (PRS), for eastern China data and demonstrated to give better results (Cao et al., 2014; Xu et al., 2010b) ...”

- p.5, line 14: I don't understand, you also use MAT? Why? Please explain more and give more details on the methods

We used MAT because relationship between pollen and HII is quite different to pollen and climate. HII is much more random and irregular comparing to climate variables, we can use it in ordination as an explanatory variable, but we cannot model HII optima or tolerance for a pollen taxon using WA or WA-PLS methods. There is no ecological basis to do like this. However, by measuring the mean analogue-HII, we have a chance to assess the potential bias induced from human impact on surface samples in reconstructing past climate. This is because the reconstructed climate of a fossil sample is mostly determined by those surface samples which have similar or close pollen spectra no matter what method is used.

We rewritten this part as: “HII was also analysed in the same way to evaluate the human impact on the pollen data and the quality of the training set. HII as an environmental variable with certain features of stochastic on locations and intensity, there is no robust ecological basis to estimate the optima and tolerance for a pollen taxon using any pollen-HII calibration set. Therefore, we used an indirect method to assess the potential bias induced from training set due to human impact on surface samples. At first, we found five closest modern analogues for each fossil sample using MAT (Simpson, 2007), and then used mean HII value at the analogue sites to examine the human influence on the analogue samples and to evaluate the bias in climate reconstruction for that given fossil sample.”

- p.5, line 21: “the differences between 151(147) and 99(93) taxa in the Natural (human) set are only explained by the exclusion of rare taxa?

It can be explained like this. There is no need to omit the rare taxa from ecological perspective, and we do this for reducing the “noise” which refers the component of variation arising from some random effects, such as pollen dispersal and deposition, analytical and counting errors, taxonomy discordance in data-sets merging, which all potentially increase the noise in a training set. With large-size and species-rich data, WA-PLS may fail to improve on WA due to complex noise (Juggins and Birks, 2012). Excluding some “rare data” is necessary in reconstructing work.

- p.5, line 38: I don't understand why you don't keep the 3 component model (even given the threshold of 5%).

We used threshold of 5% to select the ‘minimal adequate model’ to prevent over-fitting (Birks, 1998). The reduction in prediction error sometimes could be caused by fitting improvement of only a few samples, and model with more components is not a guarantee to increase the predictive power, sometimes even poorer (Juggins and Birks, 2012). In our N-set, reduction in RMSEP using 3 component model is only 1.48%, and it is not significant ($p=0.098$) in randomisation t-test. Two-component is more appropriate

in this case.

We also added the p-value (presented in Table 2) in the text: “and not significantly different ($P = 0.098$)”.

6. Discussion

- P.6, line 31 is a function of climate: yes but also other factors play a role

Agreed, we mentioned this in the next sentence: “...this indirect pollen-climate relationship can be affected by several other (non-climatic) factors”. We have rewritten this sentence as “The pollen record is a complex and nonlinear function of vegetation which in turn is a function of climate based on some key assumptions”.

- p.7, line 18: a significant bias: do you mean statistically significant? it has been tested?

We mean the reconstructed P_{ann} values of early and middle Holocene using H-set are less than N-set and the difference is statistically significant ($P < 0.001$).

We have rewritten it as: Using the H-set in this study, significantly lower ($P < 0.001$) P_{ann} values relative to the N-set are reconstructed for Lake Gonghai during the early and middle Holocene (TW3 and TW4), when tree pollen contributed more than 40% to the total pollen sum. This suggests that drier biases may have also been induced from surface samples using the N-set for this period.

- p.8, line 16: how is calculated the “mean analogue-HII values in the N-set”?

We used MAT method to find 5 best analogues in the both training sets for each fossil sample and calculated their mean HII values. We have added this information in the rewritten section 3.4.

- p.8 line 32: “This raises the question of whether the Holocene climate could be quantitatively reconstructed using pollen data from eastern China. The answer is not a simple ‘yes’ or ‘no’”; I think that the question is “when” and “where” given that your reconstruction before 2.8 ka can be considered as robust and climate driven

This is indeed a very good and important comment. Farming practice in some places within study area was quite early, however, getting a sequence with less human disturbance before 2.8 ka is still possible, such as Gonghai. Indeed, the calibration set we called natural has been also effected by human impact from both the present and the past. The deviation between reconstructed P_{ann} results is due to the human-induced bias on optima estimate in the H-set. We know that reference samples in the H-set are strongly affected by human activities. At least we can assume that the P_{ann} reconstruction based on the N-set should be closer to the ‘reality’ in the past.

7. Tables and figures

- Fig 1: a, b, the legend is not clear (the right one); the site is not easy to find on the fig: please correct

Corrected. We have changed the font size in the legend box and the site symbol and its color (see below).

References:

- Birks, H. J. B.: Numerical tools in palaeolimnology - Progress, potentialities, and problems, *J. Paleolimnol.*, 20, 307–332, 1998.
- Chen, F., Xu, Q., Chen, J., Birks, H. J., Liu, J., Zhang, S., Jin, L., An, C., Telford, R. J., Cao, X., Wang, Z., Zhang, X., Selvaraj, K., Lu, H., Li, Y., Zheng, Z., Wang, H., Zhou, A., Dong, G., Zhang, J., Huang, X., Bloemendal, J., and Rao, Z.: East Asian summer monsoon precipitation variability since the last deglaciation, *Sci. Rep.*, 5, 11186, 2015.
- 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, 1987.
- Grimm, E. C.: Tilia 1.7. 16. In: Illinois State Museum, Research and Collection Center, Springfield, 2011.
- Juggins, S. and Birks, H. J. B.: Quantitative environmental reconstructions from biological data. In: *Tracking Environmental Change Using Lake Sediments*, vol 5: Data Handling and Numerical Techniques, Birks, J. B. H., Lotter, A. F., Juggins, S., and Smol, J. P. (Eds.), Springer Netherlands, Dordrecht, 2012.
- Li, Y., Bunting, M. J., Xu, Q., Jiang, S., Ding, W., and Hun, L.: Pollen–vegetation–climate relationships in some desert and desert-steppe communities in northern China, *Holocene*, 21, 997–1010, 2011.
- Xu, Q., Chen, F., Zhang, S., Cao, X., Li, J., Li, Y., Li, M., Chen, J., Liu, J., and Wang, Z.: Vegetation succession and East Asian Summer Monsoon Changes since the last deglaciation inferred from high-resolution pollen record in Gonghai Lake, Shanxi Province, China, *Holocene*, 27, 835–846, 2016.
- Xu, Q., Li, Y., Bunting, M. J., Tian, F., and Liu, J.: The effects of training set selection on the relationship between pollen assemblages and climate parameters: Implications for reconstructing past climate, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 289, 123–133, 2010.
- Zhao, M. and Running, S. W.: Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, *Science*, 329, 940–943, 2010.

New and modified figures:

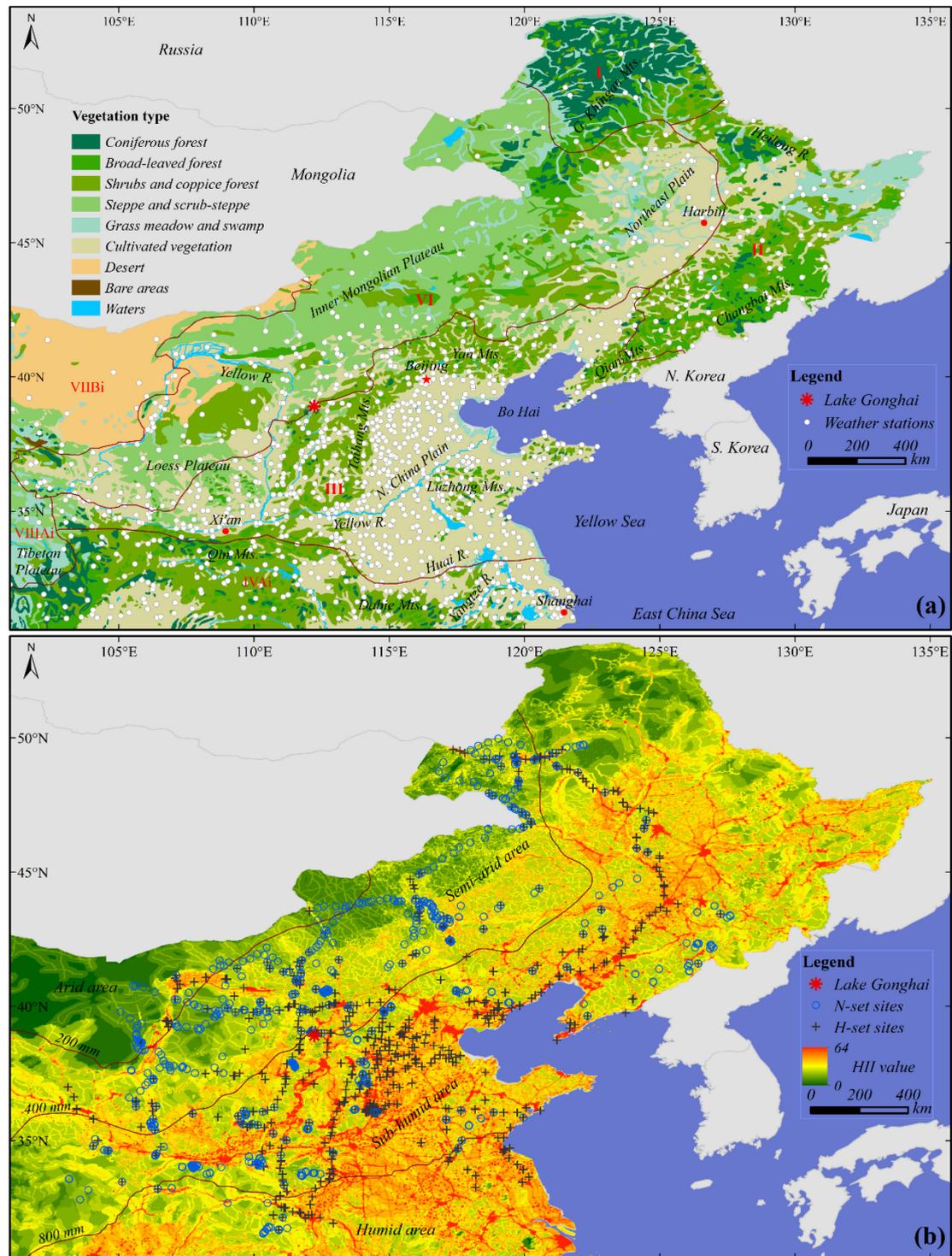


Figure 1. Maps of the study region showing the distributions of (a) vegetation regions and types, and the 1208 meteorological stations, and (b) HII values with surface pollen sampling sites from N-set (circles) and H-set (crosses). Selected large cities (red pentagram and dots), mountains (Mts.), rivers (R.), isohyets of 200, 400 and 800 mm (for annual precipitation), and location of Lake Gonghai are marked.

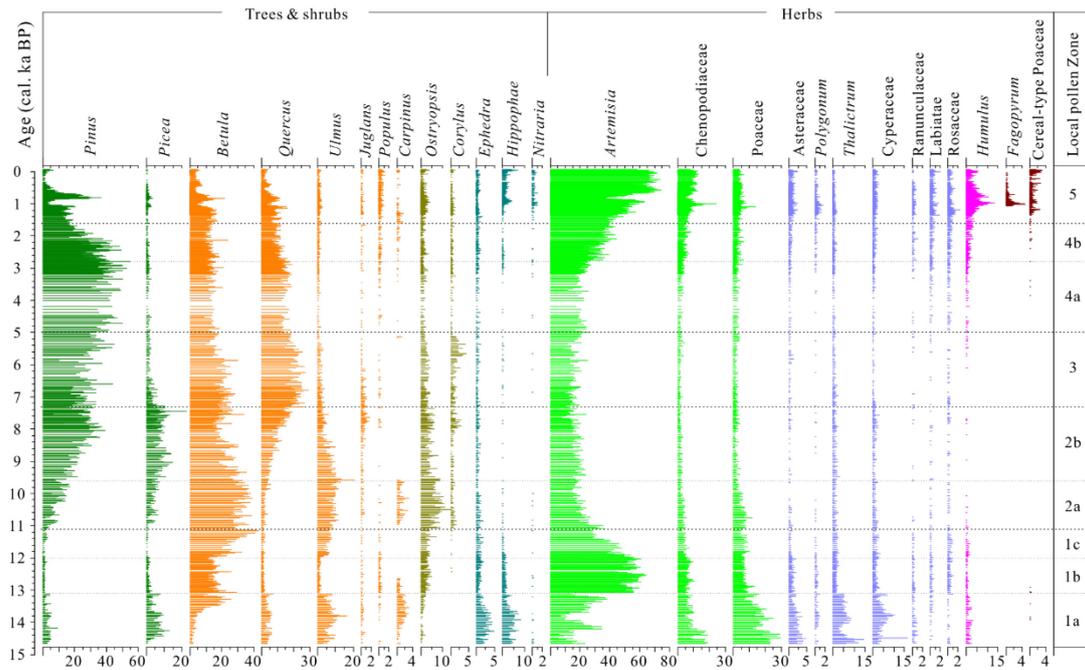


Figure 2. Simplified pollen percentage diagram of core GH09B from Lake Gonghai. The local pollen zones (and subzones) boundaries are based on the results of constrained cluster analysis CONISS in TILIA software (Grimm, 1987, 2011) used for making the pollen diagram. The detailed information of vegetation succession was presented in Xu et al. (2016).

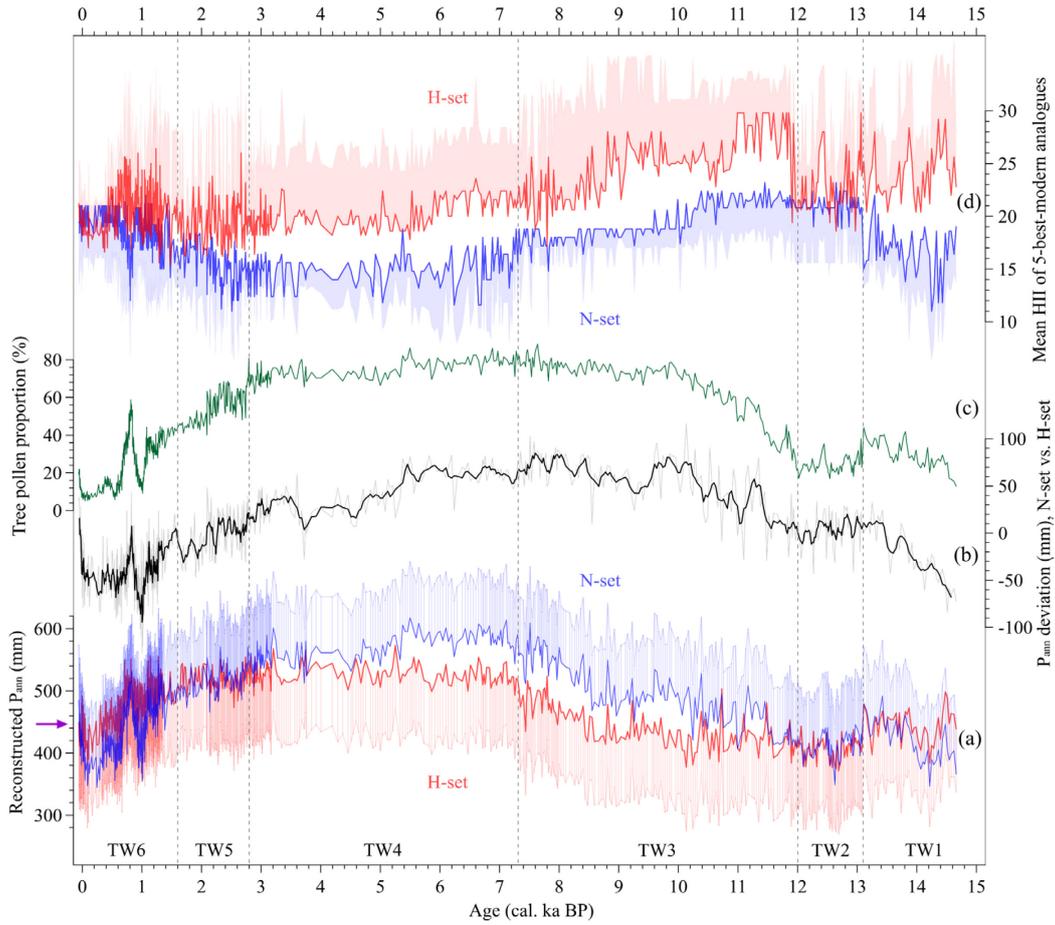


Figure 6. Fossil pollen record of Lake Gonghai: (a) reconstructed annual precipitation using a natural (N-) set (blue) with upper side standard error of prediction (SEP) and a human-induced (H-) set (red) with lower side SEP. The modern instrumental value is marked on the scale axis for comparison (purple arrow). (b) Deviations in the reconstructed precipitation (grey) between the N-set and H-set based transfer functions with a 5-point moving average smoother (black). (c) The proportion of tree pollen taxa (%), and (d) the mean HII value of the five best analogues in the N-set (blue, with lower side standard deviation) and the H-set (red, with higher side standard deviation) for fossil samples. Six time windows (TWs), delineated according to the deviation pattern between the two reconstructions, are separated by grey dashed lines.