

# Decoupled Asian monsoon intensity and precipitation during glacial-interglacial transitions on the Chinese Loess Plateau

Received: 30 July 2021

Accepted: 30 August 2022

Published online: 14 September 2022

 Check for updatesYukun Zheng<sup>1</sup>, Hongyan Liu<sup>1</sup>✉, Huan Yang<sup>2</sup>✉, Hongya Wang<sup>1</sup>,  
Wenjie Zhao<sup>1</sup>, Zeyu Zhang<sup>1</sup>, Miao Huang<sup>3</sup> & Weihang Liu<sup>4</sup>

The discrepancies among the variations in global ice volume, cave stalagmite  $\delta^{18}\text{O}$  and rainfall reconstructed by cosmogenic  $^{10}\text{Be}$  tremendously restrain our understanding of the evolution of the East Asian summer monsoon (EASM). Here, we present a 430-ka EASM mean annual precipitation record on the Chinese Loess Plateau obtained using branched glycerol dialkyl glycerol tetraethers based on a deep learning neural network; this rainfall record corresponds well with cave-derived  $\delta^{18}\text{O}$  data from southern China but differs from precipitation reconstructed by  $^{10}\text{Be}$ . Both branched tetraether membrane lipids and cave  $\delta^{18}\text{O}$  may be affected by soil moisture and atmospheric temperature when glacial and interglacial conditions alternated and were thus decoupled from atmospheric precipitation; instead, they represent variations in the intensity of the EASM. Furthermore, we demonstrate that the brGDGT-DLNN method can significantly extend the temporal scale record of the EASM and is not restricted by geographic location compared with stalagmite records.

The loess–paleosol sequences (LPSs) on the Chinese Loess Plateau (CLP), recognized by Heller & Liu<sup>1</sup>, contain a significant climatic archive and precise chronological controls for terrestrial paleoclimatic research. The chronology of loess records spanning the past three glacial–interglacial cycles in China has been dramatically improved in current years by using newly developed K-feldspar luminescence dating<sup>2</sup>. In recent decades, records of the East Asian summer monsoon (EASM) in LPSs and speleothems have considerably enhanced our comprehension of its evolutionary history and dynamic mechanisms<sup>3,4</sup>. It has been consistently accepted that the Chinese LPSs correlates well with benthic foraminiferal  $\delta^{18}\text{O}$  records, which can reflect the variations in sea level and ice-volume cycles<sup>5</sup>. The EASM reconstructed by LPSs corresponds well with the high-latitude ice volume in the Northern Hemisphere, which shows a 100-ka dominant

cycle after the Middle Pleistocene<sup>6,7</sup>. Although several proxies based on LPSs indicate concomitant 100-ka and 23-ka periods (i.e., loess  $^{10}\text{Be}$ <sup>8</sup> and  $\delta^{13}\text{C}_{\text{IC}}$ <sup>9</sup>), a few new proxies display no 23-ka cycle (i.e., microcodium Sr/Ca<sup>10</sup> and East China Sea local seawater  $\delta^{18}\text{O}$ <sup>11</sup>). These factors all restrain our understanding of the evolution of the EASM and, more importantly, its response to ice-volume forcing and insolation. Another significant proxy for the EASM, stable oxygen isotopes of speleothems in caves from southern China with robust chronological controls<sup>12</sup>, displays a dominant precessional period (23 ka), which indicates that the EASM is directly controlled by summer insolation in the Northern Hemisphere<sup>13,14</sup>. However, the interpretations of speleothem  $\delta^{18}\text{O}$  are controversial<sup>15</sup> due to, for example, regional rainfall, the seasonality of precipitation over the cave site<sup>4</sup>, and the outwind rainout of air masses<sup>16</sup>.

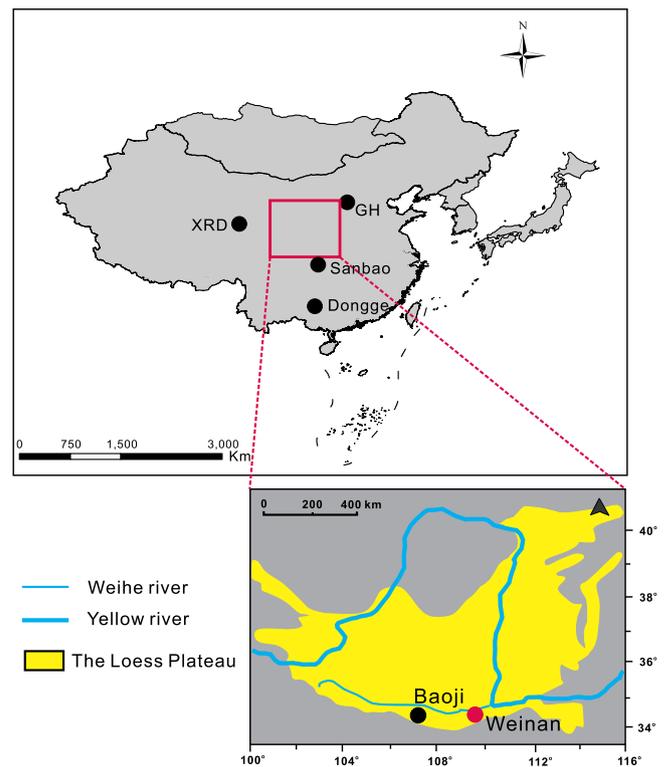
<sup>1</sup>College of Urban and Environmental Sciences and MOE Laboratory for Earth Surface Processes, Peking University, Beijing, China. <sup>2</sup>Hubei Key Laboratory of Critical Zone Evolution, School of Geography and Information Engineering, China University of Geosciences, 430074 Wuhan, China. <sup>3</sup>State Key Laboratory of Biogeology and Environmental Geology, School of Earth Sciences, China University of Geosciences, 430074 Wuhan, China. <sup>4</sup>Academy of Disaster Reduction and Emergency Management, Faculty of Geographical Science, Beijing Normal University, 100875 Beijing, China.

✉ e-mail: [lhy@urban.pku.edu.cn](mailto:lhy@urban.pku.edu.cn); [yhsailing@163.com](mailto:yhsailing@163.com)

Some studies have attempted to reconstruct the paleoclimate, including the mean annual temperature (MAT) and mean annual precipitation (MAP), on the CLP. Phytoliths were used in the early years of research as a proxy to reconstruct MAT and MAP on the CLP<sup>17,18</sup>. However, the low classification resolution of phytoliths may expand their tolerance to climatic factors and affect their sensitivity to environmental variations<sup>17</sup>. Another proxy, the carbon isotopic composition of bulk soil organic matter ( $\delta^{13}\text{C}_{\text{SOM}}$ ), reflects the relative abundances of  $\text{C}_3$  and  $\text{C}_4$  plants for reconstructing MAP on the CLP<sup>19</sup>. However, this proxy is also affected by temperature, making it difficult to tease apart the precipitation signal<sup>20</sup>. Afterward, the utilization of cosmogenic  $^{10}\text{Be}$ , which is generated in the atmosphere and readily adsorbed by aerosols and then carried to surface sediments with precipitation, made it possible to quantitatively reconstruct MAP on the CLP, revealing the low-latitude interhemispheric insolation-driven EASM intensity<sup>8</sup>.

One promising geochemical proxy, branched glycerol dialkyl glycerol tetraethers (brGDGTs) (Supplementary Fig. 1), was confirmed to have a bacterial origin<sup>21</sup>, and a vast number of studies have investigated the environmental controls on this proxy. The study sites where brGDGTs have been used cover a great part of the world, and brGDGTs lipids can persist in the environment up to at least 55 million years ago<sup>22</sup>. Previous empirical observations have revealed that the methylation and cyclization degree of brGDGTs have tight connections with MAT and soil pH, respectively<sup>23–26</sup>. More specifically, many studies have reconstructed paleotemperature and paleosol pH changes at various temporal and spatial scales globally based on these relationships<sup>27–29</sup>. Three structural varieties of brGDGTs, containing an uncharacterized isomer of brGDGT Ic (Supplementary Fig. 1), have been confirmed in an Acidobacteria species that was cultured under oxygen limitation<sup>30</sup>. Nevertheless, it has been difficult to trace the exact bacterial sources of the other brGDGTs, and subsequent studies have revealed that brGDGTs may be derived from aquatic environments<sup>31</sup>. Recently, some studies have attempted to use various statistical approaches to improve the correlation coefficients between brGDGTs proxies and climate data by setting up nonlinear relationships. For example, Crampton-Flood and coauthors utilized the Bayes model to enhance the precision of MAT estimation based on brGDGTs<sup>32</sup>. Moreover, Martinez-Sosa et al.<sup>33</sup> and Ragberg et al.<sup>34</sup> also used nonlinear approaches to calibrate brGDGTs in lakes to temperature, enhancing the precision of models in predicting terrestrial surface temperatures. In addition to temperature, other factors, MAP, soil water content and bacterial community structure, were found to affect brGDGTs based proxies<sup>35</sup>. Although brGDGTs definitely have bacterial origins, empirical studies have found that the soil water content (SWC) may affect the methylation degree of brGDGTs, and 6-methyl brGDGTs isomers are significantly affected by SWC, especially in arid and semiarid regions<sup>36</sup>. When the relative amount of 6-methyl vs. 5-methyl brGDGTs ( $\text{IR}_{6\text{ME}} > 0.5$ ), one promising index, MBT', which expresses the methylation degree of brGDGTs, can potentially represent variations in SWC. However, it is unprecedented for brGDGTs to forecast MAP based on established quantitative transfer functions. In addition, the temperature reconstructed by brGDGTs on the CLP does not correspond well with glacial–interglacial cycles<sup>27</sup>, and discrepancies are also observed between the rainfall amounts reconstructed by  $^{10}\text{Be}$  and Chinese cave  $\delta^{18}\text{O}$  values<sup>8</sup>.

In this study, we established an optimal MAP model based on all 6-methyl brGDGTs, Ia and Ic, and applied this method to a well-dated loess–paleosol profile in Weinan (430 ka before present (BP), Fig. 1) covering five glacial–interglacial periods. Combined with global terrestrial brGDGT-calibrated temperature based on some of the 5-methyl brGDGTs isomers and soil pH models based on some of the 6-methyl brGDGTs isomers established by De Jonge and colleagues<sup>24</sup>, we reconstructed the variations in MAT and paleo-pH after 430 ka BP in Weinan profile. Furthermore, using a promising SWC proxy, MBT', we



**Fig. 1 | Geographic locations of Weinan and Baoji loess–paleosol profiles on the Chinese Loess Plateau.** And the locations of study sites mentioned in this text include: Sanbao cave<sup>8</sup>, Xiangride (XRD)<sup>55</sup>, Gonghai (GH)<sup>56</sup>, and Dongge cave<sup>4</sup>. The base map was generated using ArcGIS software.

obtained the relative SWC changes in the derived profile. We hypothesized that the brGDGT-based mean annual precipitation (MAPc) might deviate from the actual atmospheric precipitation, different from that reconstructed by  $^{10}\text{Be}$  from the atmosphere. Instead, the records of brGDGTs may indeed reveal a complicated variation in the EASM that contains the temperature or SWC signal.

## Results

### brGDGT-MAP models

Here, we used 712 previously published brGDGTs surface soil samples from around the world (Supplementary Figs. 2 and 3) and indicated the distributions of various brGDGTs compounds. In addition, the MAP data of each soil site applied in this study were all published in previous studies (Table 1). Through analyses of the linear relationships between the fractional abundances (FAs) and MAP obtained from various brGDGTs compounds, we found that only a few brGDGTs compounds correlated well with MAP: Ia (Supplementary Fig. 4a) was 0.55, Ia' (Supplementary Fig. 4e) was 0.34, and IIIa' (Supplementary Fig. 4k) was 0.30. The majority of brGDGTs isomers had low  $R^2$  values, indicating that they have little direct correlation with MAP. In particular, the  $R^2$  values for the correlations between the FAs of Ib, IIa, IIc, and MAP were  $4.56 \times 10^{-4}$ , 0.003, and 0.007, respectively (Supplementary Fig. 4b, d, h).

We then selected 533 surface soil samples as the training dataset and 179 surface soil samples as the validation dataset, both of which satisfied the principle of randomness. We assessed the precision of the model using the values of the validation dataset's  $R^2$  and root mean square error (RMSE) values. By means of several parameter tests in brGDGT-MAP models (Supplementary Figs. 5–9), our deep-learning neural network (DLNN) methods could precisely capture the multi-dimensional and nonlinear relationships between the brGDGTs isomers and MAP Supplementary Fig. 10, for training dataset:  $R^2 = 0.82$ ,

**Table 1 | List of global surface soil datasets**

Reference	Quantity of surface soils	brGDGTs compound types	DOI
De Jonge et al. <sup>24</sup>	191	15	<a href="https://doi.org/10.1016/j.gca.2014.06.013">https://doi.org/10.1016/j.gca.2014.06.013</a>
Xiao et al. <sup>58</sup>	27	15	<a href="https://doi.org/10.1016/j.orggeochem.2015.10.005">https://doi.org/10.1016/j.orggeochem.2015.10.005</a>
Yang et al. <sup>59</sup>	26	15	<a href="https://doi.org/10.1016/j.orggeochem.2015.02.003">https://doi.org/10.1016/j.orggeochem.2015.02.003</a>
Lei et al. <sup>60</sup>	44	15	<a href="https://doi.org/10.1016/j.orggeochem.2016.02.003">https://doi.org/10.1016/j.orggeochem.2016.02.003</a>
Wang et al. <sup>61</sup>	93	15	<a href="https://doi.org/10.1016/j.quascirev.2020.106172">https://doi.org/10.1016/j.quascirev.2020.106172</a>
Wang et al. <sup>62</sup>	56	15	<a href="https://doi.org/10.1016/j.orggeochem.2016.05.013">https://doi.org/10.1016/j.orggeochem.2016.05.013</a>
Ding et al. <sup>63</sup>	27	15	<a href="https://doi.org/10.5194/bg-12-3141-2015">https://doi.org/10.5194/bg-12-3141-2015</a>
Naafs et al. <sup>23</sup>	96	15	<a href="https://doi.org/10.1016/j.orggeochem.2017.01.009">https://doi.org/10.1016/j.orggeochem.2017.01.009</a>
Crampton-Flood et al. <sup>32</sup>	77	15	<a href="https://doi.org/10.1016/j.gca.2019.09.043">https://doi.org/10.1016/j.gca.2019.09.043</a>
Wang et al. <sup>64</sup>	28	15	<a href="https://doi.org/10.1038/s41598-017-17964-0">https://doi.org/10.1038/s41598-017-17964-0</a>
Wang et al. <sup>65</sup>	47	15	<a href="https://doi.org/10.1038/s41598-019-39147-9">https://doi.org/10.1038/s41598-019-39147-9</a>

RMSE = 291 mm, and  $n = 533$ ; for the validation dataset:  $R^2 = 0.80$ , RMSE = 287 mm, and  $n = 179$ ). As a result, brGDGTs can serve as optimal tools for reconstructing global MAP variations, especially considering their global ubiquity.

### Paleoclimate and paleoenvironment reconstruction in Weinan

In combination with the age-depth model obtained by interpolation between geomagnetic polarity boundaries and using magnetic susceptibility as an indicator of accumulation rate, as well as the  $U-^{230}Th$ -dated oxygen isotope records from Sanbao caves in central China (Supplementary Fig. 15), paleotemperature records reconstructed from brGDGTs in the Weinan loess–paleosol profile after 430 ka BP (Fig. 2a) could aid in further interpreting the timing of glacial terminations in northern China. Indeed, based on multiyear observed temperature records at the nearest meteorological station, the modern mean annual temperature in Weinan was 13.8 °C, which is similar to the brGDGT-inferred temperature at the top of  $S_0$ . Moreover, our 430-ka MAT records clearly show the dominance of the 100-kyr cycles, which correspond well with the loess–paleosol cycles and reveal the variations (in a range of ~10 °C) in MAT on glacial–interglacial timescales. The magnetic susceptibility and mean grain-size (MGS) records also display a majority cycle of 100 ka (Fig. 3b).

Regarding the soil paleo-pH, the temperature changes were almost opposite to the variations in soil pH over the last 430 ka BP (Fig. 2a). Similar to the MAT reconstruction, the dominant soil pH cycle also showed a 100-kyr period and revealed apparent variations (within a range of ~0.5) in pH on glacial–interglacial timescales. We found that the pH value in  $S_4$  reached its lowest level at 430 ka, and the soil was acidic for a long time. The results showed that the minimum and maximum pH values recorded over the last 430 ka were 6.1 (~380 ka) and 7.6 (~0.2 ka), respectively. The most extensive pH range in the five interglacial periods ( $S_0$ – $S_4$ ) was ~1.1 that was found in  $S_4$ .

To obtain the MAP reconstructions, we entered all 6-methyl brGDGTs, Ia and Ic samples into the brGDGT-MAP model, which may be suitable for reconstructions on the CLP, especially in semiarid and semihumid regions. As a result, we obtained the variations in MAPc in Weinan over the past 430 ka (Fig. 2c). We found an obvious glacial–interglacial cycle in this MAPc curve. The absolute value of MAPc in the Weinan profile was almost equal to the observed multiyear MAP in this region (2000–2015) when the samples approach the surface ( $S_0$ ).

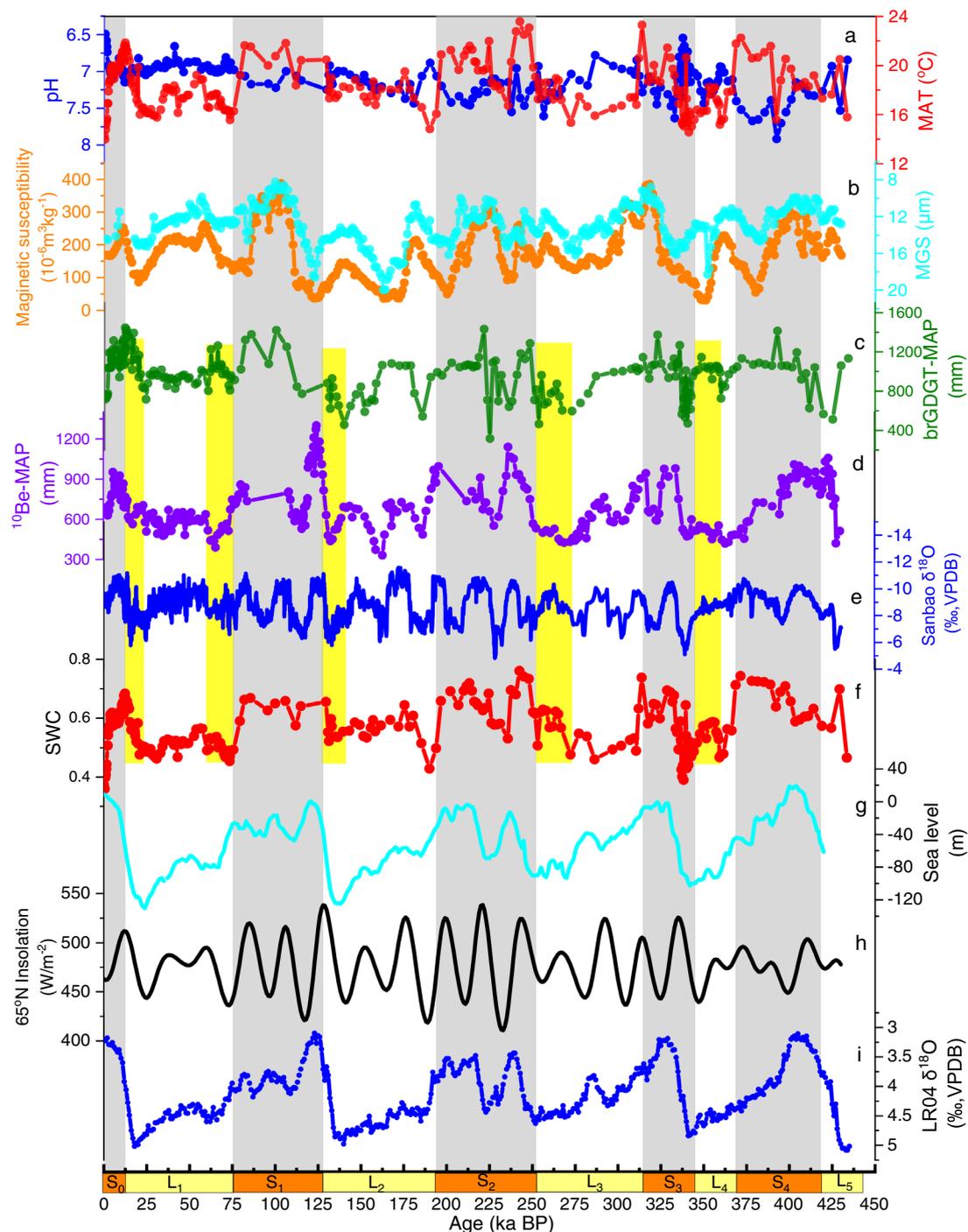
With regard to SWC, we found that nearly all  $IR_{6ME}$  values, exceeded 0.5 in this profile (Supplementary Fig. 16). As a result, MBT' can be regarded as a potential SWC indicator in the Weinan LPS. Here, we found a dominant 100-kyr cycle in the MBT' record (Fig. 2f), representing relatively high SWC during interglacial periods and relatively low SWC values in glacial periods in 430 ka BP in the Weinan LPS.

### Discussion

With the brGDGT-MAP model we established and the paleoclimate we reconstructed in the Weinan profile over the past 430 ka, we can better understand the variation in EASM history in this region. Regarding the brGDGT-MAP model, we found that the 5-methyl and 6-methyl brGDGTs isomers display different linear correlations with MAP, and this difference may be due to the 5-methyl brGDGTs isomers containing a more significant temperature signal compared with the pH, soil moisture and MAP signals<sup>24</sup>. On the other hand, some studies have revealed that 6-methyl brGDGTs are more responsive to soil pH and may be more affected by soil moisture than other indicators, and these variables have close relationships with MAP<sup>36</sup>. More importantly, we reveal that the relationships between brGDGTs and modern MAP are not direct linear correlations. The adequate evidence we conclude is as follows: first, we entered the same arguments (all 6-methyl brGDGTs isomers, Ia and Ic) into the multiple linear regression model, and the results show the poor forecast ability both in the training dataset and validation dataset, especially when those soil-sites measured MAP > 1500 mm (Supplementary Figs. 5 and 6). Furthermore, we indicate that the larger linear  $R^2$  value between the 6-methyl isomer and MAP does not indicate a better capacity of this indicator to predict MAP. For example, although the linear  $R^2$  value of Ic and MAP is obviously larger than that of IIIc' and MAP (Supplementary Fig. 4), the IIIc' isomers are more critical than Ic in predicting the MAP values in our DLNN model (Supplementary Fig. 9). In addition to Ic and IIIc' in our brGDGT-MAP model, the  $R^2$  values of the validation dataset decrease more when excluding IIIc' than when excluding Ic. We further utilized a more complex artificial neural network (ANN) (i.e., recurrent neuron network (RNN), long short-term memory networks (LSTM) and gate recurrent unit (GRU)) to predict MAP based on brGDGTs (Supplementary Fig. 11). The results also reveal that our DLNN model is the most suitable one to apply in this field (Supplementary Fig. 12). Finally, we tested this brGDGT-MAP model in the Xiangride (XRD) profile (Fig. 1), and the variations in the Holocene MAP we reconstructed correspond well with the most reliable rainfall proxies in the EASM region (Supplementary Fig. 14) and we concluded that brGDGTs can be used as a robust tool to quantitatively reconstruct MAP on a suborbital time scale.

A previous study found a robust relationship between water balance and soil pH<sup>37</sup>. The latter can be calculated well by 6-methyl brGDGTs. More recently, it was found that the distribution of Actinobacteria is positively correlated with aridity, maintaining the same trend as that of 6-methyl brGDGTs<sup>38</sup>. Indeed, our study further revealed the quantitative relationship between MAP and 6-methyl isomers through the DLNN model.

The MAPc we reconstructed differs from the reconstruction obtained from  $^{10}Be$ . Although quite a few studies have developed

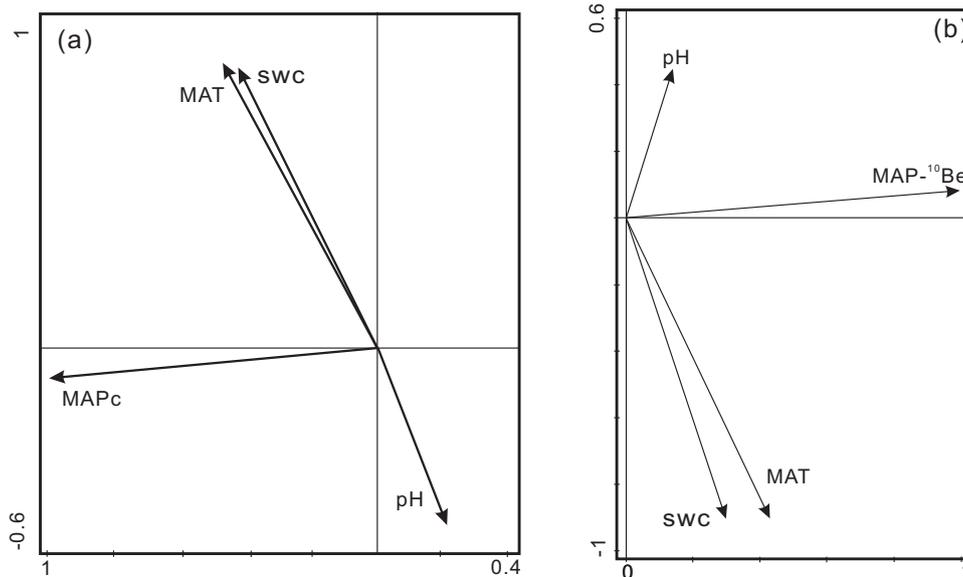


**Fig. 2 | Variations in brGDGTs parameters compared with the parameters of other records.** **a** The pH calibration based on 5- and 6-methyl branched glycerol dialkyl glycerol tetraethers (brGDGTs, dark blue solid line, this study) and mean annual average temperature (MAT) calibration based on 5- and 6-methyl brGDGTs (dark red solid line, this study). **b** The mean grain size (mm) (MGS, blue solid line, this study) and magnetic susceptibility (orange solid line, this study). **c** The mean annual average precipitation (MAPc) values reconstructed by the brGDGT-MAP model (Weinan, dark green solid line, this study). **d** The MAP values reconstructed by  $^{10}\text{Be}$  (Baoji, purple solid line). **e** The variations in Sanbao  $\delta^{18}\text{O}$  (the dark blue solid

line indicates the results of the five-point average). **f** The changes in the MBT' (dark red solid line), representing the soil water content (SWC) variations<sup>36</sup>. **g** The changes in sea level (blue solid line)<sup>37</sup>. **h** The solar radiation at 65°N (black solid line)<sup>42</sup>. **i** The benthic foraminifera-derived  $\delta^{18}\text{O}$  stack (dark blue solid line)<sup>41</sup>. The orange and yellow bars represent the loess–paleosol cycles in the Weinan profile. The gray shading bars represent the interglacial periods in the Weinan profile, and the light-yellow bars highlight the decoupled periods between the  $^{10}\text{Be}$  reconstructions and our results.

quantitative models between brGDGTs indices and MAT or soil pH values on global or regional scales, MAP is difficult to reconstruct in the Weinan LPS. Compared with reconstructions of temperature and soil pH, MAP reconstructions, especially on geologic timescales, are more difficult to obtain and lack reliable proxies. Through the brGDGT-MAP

model, we obtained the MAPc changes over the past 430-ka BP (Supplementary Fig. 15)<sup>39,40</sup>. Another MAP reconstruction was obtained by an empirical cosmogenic proxy ( $^{10}\text{Be}$ ) in Baoji (CLP, China, Figs. 1 and 2d), located in northwestern Weinan. In terms of the modern MAP observations and geographic distribution pattern, the MAP values we



**Fig. 3 | Principal component analysis (PCA) among various environmental factors on the Chinese Loess Plateau. a** brGDGT-MAP (mean annual average precipitation), MAT (mean annual average temperature), SWC (soil water content),

and pH in Weinan since 430 ka BP ( $n = 136$ ) and **b**  $^{10}\text{Be}$ -MAP, MAT, SWC and pH when glacial and interglacial periods alternated ( $n = 77$ ). Totals of 98% (a) and 90% (b) of the variations in the paleoclimate data were explained by the first two PCA axes.

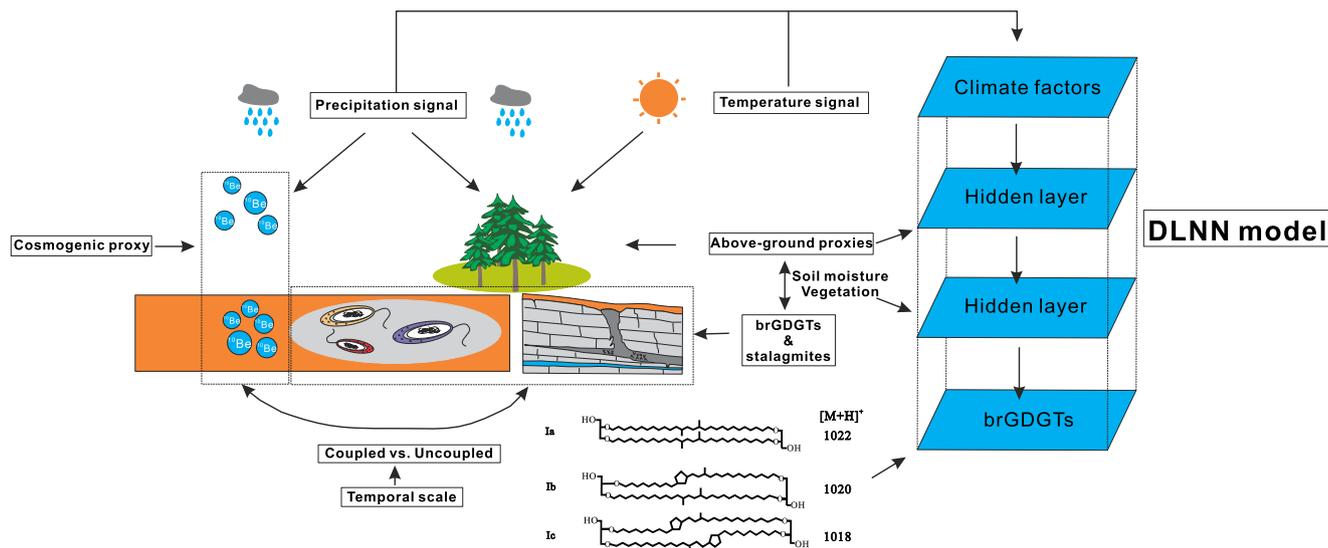
obtained in Weinan are close to those obtained in Baoji. Indeed, our MAP reconstructions slightly exceed the MAP values reconstructed by  $^{10}\text{Be}$  in Baoji, spanning almost the whole 430 ka. Moreover, the MAP changes observed between the glacial and interglacial periods are nearly concentrated within similar ranges when the brGDGT-MAP model and  $^{10}\text{Be}$  are used. However, we revealed that opposite variations existed between the MAP values reconstructed by  $^{10}\text{Be}$  and MAPc values we reconstructed by using brGDGTs when the glacial and interglacial conditions were altered (Fig. 2, light-yellow bars). The  $^{10}\text{Be}$ -derived MAP is weakly correlated with the Chinese cave-derived  $\delta^{18}\text{O}$  (Fig. 2e) but aligns with global ice-volume variations (Fig. 2g, i)<sup>41</sup>, which may be affected by 65°N summer insolation in the Milankovitch hypothesis<sup>42</sup>. Thus, previous authors have argued that Chinese cave-derived  $\delta^{18}\text{O}$  is governed by low-latitude insolation rather than high-northern-latitude insolation (Fig. 2h).

Nevertheless, our reconstructed MAPc values corresponded well with the Chinese cave-derived  $\delta^{18}\text{O}$ , which is a significant proxy of EASM intensity and is also likely affected by precipitation, soil moisture and transportation paths<sup>43</sup>. Our MAPc results do not match the SWC and MAT variations that occurred in the last 430 ka (Fig. 3a). We found that 23-ka cycles dominated MAPc; these results do not correspond with the 100-ka cycles observed in SWC and MAT (Fig. 2a, c, f). A previous study also found a close and consistent relationship between MAT and SWC at glacial–interglacial scales<sup>44,45</sup>. In addition, we found that our MAPc results had relatively weak relationships with MAT and SWC (Fig. 3a). The reconstructed MAT and potential SWC (MBT) values showed strong positive correlations with each other, and they were both negatively connected with soil pH. Concerning the co-vary relationship between MAT and SWC that might potentially be affected by the calibration model's parameters we used, we have tested the different temperature (four models) and SWC (two models) calibration models, which show the same variation trends between all MAT and SWC reconstructions (Supplementary Fig. 17). These relationships are similar to the relationships between  $^{10}\text{Be}$ -MAP and the three paleoclimatic and paleoenvironmental factors we reconstructed (MAT, soil pH, and SWC) (Fig. 3b). As a result, the real atmospheric MAP might have a weak connection with the reconstructed MAT and SWC on glacial–interglacial temporal scales. Furthermore, combined with other studies and our results<sup>45</sup>, we concluded that the environmental

SWC variations in the CLP are more affected by MAT than by MAP, due to the greater water evaporation and sparse vegetation cover in this area.

The isotope  $^{10}\text{Be}$ , as a cosmogenic proxy, may more directly record atmospheric precipitation, while belowground proxies (i.e., brGDGTs and stalagmite  $\delta^{18}\text{O}$ ) may be affected by surface variations<sup>27,46</sup> (i.e., vegetation and SWC, Fig. 4). Compared with the promising SWC proxy, MBT (Fig. 2f), which can be applied in paleoclimate reconstructions when  $\text{IR}_{6\text{ME}} > 0.5$  (Supplementary Fig. 16)<sup>24,36</sup>, we reveal that the uncoupled sections between  $^{10}\text{Be}$ -MAP and our results may be more strongly affected by SWC and MAT.  $^{10}\text{Be}$ -MAP is weakly connected with brGDGT-MAP when glacial and interglacial periods alternate, and this relationship becomes a positive connection when these periods are eliminated (Supplementary Fig. 18).

The brGDGTs and cave-derived  $\delta^{18}\text{O}$  may contain SWC and MAT signals and may only represent changes in the EASM intensity during some periods at glacial–interglacial scales. Other East Asian monsoon proxies, magnetic susceptibility and mean grain size, also show the same trends in these periods (Fig. 2b). In addition, we reconstructed relatively low soil pH values and warm environments in these periods (Fig. 2a), and the temperature variations in the Weinan profile obtained based on the phytoliths are shown to be similar to our results<sup>18</sup>. Previous studies have recorded that the climate in the continental interior had considerably diverse fluctuation forms in the glacial–interglacial cycle and even on shorter timescales<sup>47</sup>. Indeed, both SWC and temperature in monsoonal East Asia covaried with global ice volume and greenhouse gas (GHG) variations<sup>45</sup>. Combined with another vegetation-based biomarker, long-chain *n*-alkanes, which are derived from plant leaf waxes, can transport signals of changes in plant sources and past climate. The carbon preference index (CPI) and the average chain length (ACL) values revealed a relatively high SWC during the glacial and interglacial transition, we further confirmed these SWC fluctuations in long-term climate change (Supplementary Fig. 13). As a result, both brGDGTs and cave-derived  $\delta^{18}\text{O}$  may deviate from the real precipitation variations when MAP and SWC are decoupled; instead, they may denote reliable EASM proxies at glacial–interglacial scales. Indeed, the MAPc we reconstructed in Weinan shows the same primary cycle with the Sanbao cave-derived  $\delta^{18}\text{O}$  at the period of 23 ka. In contrast, during



**Fig. 4 | Schematic diagram reflecting the divergent responses between cosmogenic proxies and proxies of underground sources.** All the elements are created and hand drawn by the first author of this article (Y.Z.).

this primary period (i.e., 23 ka), they demonstrate an anti-phase variation trend (Fig. 5). Furthermore, brGDGTs have significant temporal and spatial superiority compared with stalagmite proxies, can be preserved in the environment for up to at least 55 Ma and do not have geographic restrictions<sup>22</sup>.

In summary, both brGDGTs and cave-derived  $\delta^{18}\text{O}$  may contain SWC and MAT signals and do not always maintain the same changes as those observed in MAP at glacial–interglacial scales. Although these proxies do not reflect rainfall changes at all times, they can represent long-term variations in EASM intensity. In addition, the utilization of our brGDGT-MAP model can vastly enrich EASM records in both temporally and spatially in the future.

## Methods

### Materials

Weinan city is located in the middle reaches of the Yellow River and in the southern part of the Loess Plateau (34°13′–35°52′N, 108°58′–110°35′E) (Fig. 1). It has a temperate semihumid, semiarid climate. The modern MAT observations indicate a value of 13.8 °C, and MAP is 570 mm; these values were obtained from the China meteorological data network, comprising the meteorological data of 2000–2015 (<http://data.cma.cn/>). Weinan has four distinct seasons, with hot and rainy conditions occurring in the same season. Much of the annual precipitation falls from June to August. The Weinan profile contains 42.8 m of loess–paleosol sequences (LPSs), including five paleosol layers from  $S_0$ – $S_4$  and five loess layers from  $L_1$ – $L_5$  and covering five glacial–interglacial cycles. The sampling method involved collecting one sample every 10 cm without interruption. A total of 427 samples were collected from this profile.

### Modern brGDGTs dataset and MAP dataset

Previously published brGDGTs data from surface soil samples were extracted using an established brGDGT-MAP model. The surface soil samples contain various types of soil and cover nearly all climatic and latitudinal zones. These datasets contain 712 surface soil samples, which all have separated 5-methyl and 6-methyl brGDGTs isomers (Table 1). To reduce the errors in collecting data from different laboratories, the MAP datasets we entered into the brGDGT-MAP model were all published in their previous studies, and we calculated the fractional abundances of each brGDGTs compound for each sample (Table 1), although there were no data regarding changes in soil occurring based on the brGDGTs indices among various

laboratories. To eliminate and test the error of the previous MAP dataset, in this study, we also extracted each soil site’s multiyear MAP (1990–2020) through TerraClimate, which is a dataset of high-spatial-resolution monthly climate for global terrestrial surfaces (1/24°, ~4 km)<sup>48</sup>. TerraClimate datasets reveal significant advances in the overall mean absolute error and enhance spatial realism compared with coarser resolution gridded datasets. Supplementary Fig. 3 shows that the two MAP datasets have high correlations, with only a few sites exhibiting large deviations. In this study, we entered these two MAP datasets into the DLNN model to obtain the most suitable DLNN-MAP model.

### Grain-size and magnetic susceptibility measurements

Samples at 10-cm intervals were dried in an oven at 40 °C for 3 days. Then, 0.2 g of each sample was weighed using a clean beaker with an electronic balance. Then, 10 ml of 30%  $\text{H}_2\text{O}_2$  and 10 ml 10% HCl were added to remove organic matter and carbonate, respectively. Before the grain-size measurement, 0.05 mol/L ( $\text{NaPO}_3$ )<sub>6</sub> was added, and the solutions were placed in an ultrasonic machine for 10 min. The magnetic susceptibility of the samples were measured with an MS-2B Bartington meter. The grain-size was measured using a Mastersizer 2000 produced by Marvern Company in the UK, with an error of less than 1%.

### Chronology

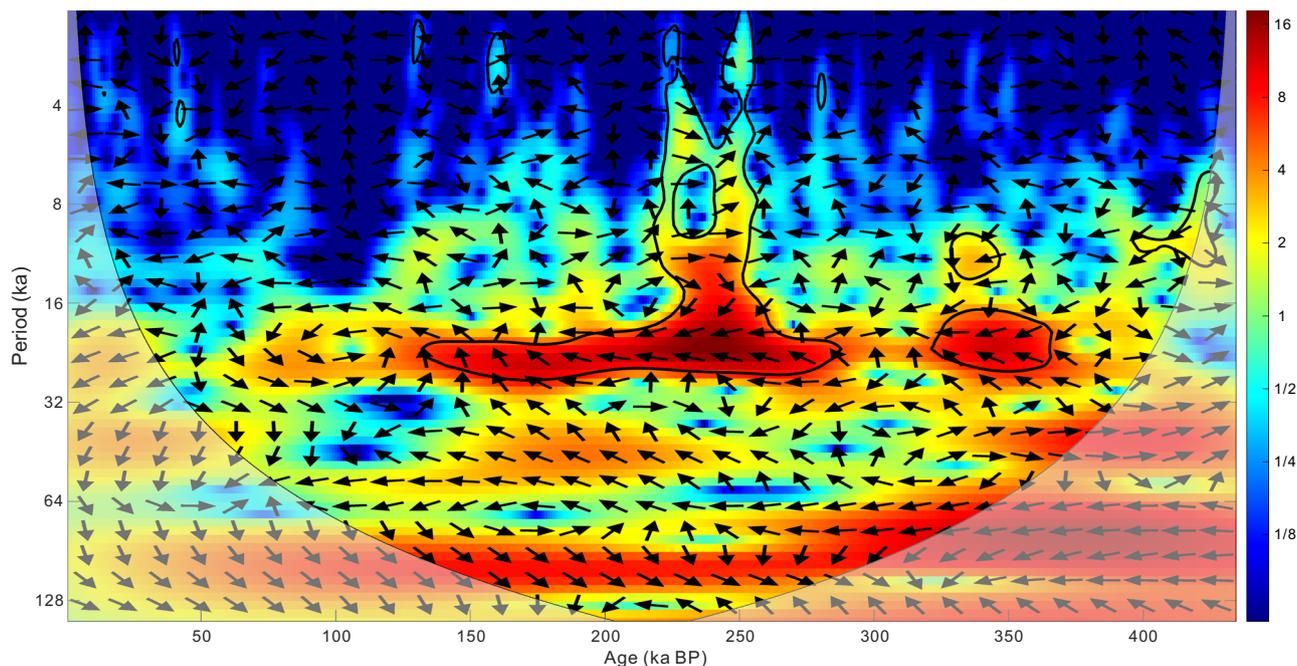
We used the ages of LPS control points on the Loess Plateau to obtain the age of each sample in the Weinan profile<sup>40</sup>. We used the magnetic susceptibility as an indicator of the accumulation rate<sup>39</sup> combined with the  $\text{U-}^{230}\text{Th}$ -dated oxygen isotope records from Sanbao caves in central China<sup>14</sup>. Each sample’s magnetic susceptibility was analyzed at 10-cm intervals (Supplementary Fig. 7). The calculation was as follows:

$$T_m = T_1 + \frac{(\sum_{i=1}^m a_i s_i)(T_2 - T_1)}{\sum_{i=1}^n a_i s_i} \quad (1)$$

where  $T_1$  and  $T_2$  indicate the ages of the control points,  $a_i$  indicates the thickness of the layer, and  $s_i$  indicates the magnetic susceptibility of the layer.

### GDGTs analysis

Lipids in a total of 238 LPS samples were extracted, including the 198 samples reported in ref. 49. Forty samples at depths from 34.9 m to



**Fig. 5 | Cross wavelet spectral analysis between high-resolution Sanbao  $\delta^{18}\text{O}$  and mean annual average precipitation (MAP) we reconstructed in Weinan.** Cross wavelet spectral output the same time resolution (1 ka) between the  $\delta^{18}\text{O}$  and

MAP though linear interpolation. The arrows that pointed to the left indicate the reverse phase change of the two variables and the bigger value indicated by different colors represents the higher relationship of the two variables.

43.7 m were selected every 20 cm intervals from the Weinan profile, and dried in an oven at 40 °C for 3 days. Afterward, the loess and paleosol samples were ground into powder and passed through a 60-mesh sieve. Each sample was weighed and extracted with 80 ml of methanol: dichloromethane (DCM) (1:9, v/v) using accelerated solvent extractors (ASE 100 or 150, Dionex, USA). The temperature and pressure were set at 100 °C and 1400 psi, respectively. Then, the lipid extracts were condensed in a rotary evaporator at 40 °C and separated into apolar and polar fractions on a flash silica gel column (0.7 cm i.d. and 1.5 g activated silica gel) chromatography using *n*-hexane and methanol as eluents, respectively. All polar components were passed through a 0.45- $\mu\text{m}$  PTFE syringe filter. All apolar and polar compositions were dried under a gentle stream of nitrogen gas.

The GDGTs were analyzed by using an Agilent 1200 series liquid chromatography-atmospheric pressure chemical ionization-6460A triple quadrupole mass spectrometry (LC-APCI-MS/MS). Ten microlitres of  $\text{C}_{46}$  GTGTs (0.001157  $\mu\text{g}/\mu\text{l}$ ) were added to each polar fraction, and the samples were then dissolved in 300  $\mu\text{l}$  of *n*-hexane: isopropanol (IPA) (98.2:1.8, v/v). Two silica gel columns in series (150 mm  $\times$  2.1 mm, 1.9  $\mu\text{m}$ , Thermo Finnigan; USA) were used for the separation of 5-methyl and 6-methyl brGDGTs, with the column temperature kept at 40 °C. The mass spectrometry settings were as follows: the vaporizer pressure 60 psi, the vaporizer temperature 400 °C, the flow rate of dry gas ( $\text{N}_2$ ) 6 l/min, drying gas temperature 200 °C, the capillary voltage 3500 V, the corona current 5  $\mu\text{A}$  ( $\sim$ 3200 V), and a single-ion monitoring mode (SIM) was used<sup>50</sup>, targeting the protonated molecular ions ( $[\text{M} + \text{H}]^+$ ) 1304, 1302, 1300, 1298, 1296, 1292, 1050, 1048, 1046, 1036, 1034, 1032, 1020, 1018, and 744.

The  $\text{MAT}_{\text{mr}}$  proxy was calculated to identify the changes that occurred in the mean annual temperature in the Weinan section over the last 430 ka. The calculation was as follows<sup>24,51</sup>.

$$\text{MAT}_{\text{mr}} = 7.17 + 17.1*[Ia] + 25.9*[Ib] + 34.4*[Ic] - 28.6*[IIa] \quad (n = 222, R^2 = 0.68, \text{RMSE} = 4.6 \text{ } ^\circ\text{C}, P < 0.01)$$

(2)

$$\text{MAT}_{\text{mr}} = 5.58 + 17.91*[Ia] - 18.77*[IIa] \quad (3)$$

$$\text{MAT}(\text{SSM}) = 20.9 - 13.4*[IIa + IIa'] - 17.2*[IIIa + IIIa'] - 17.5*[Ib + Ib'] + 11.2*[Ib] \quad (4)$$

$$\text{MAAT} = 0.81 - 5.67*\text{CBT}' + 31.0*\text{MBT}' \quad (5)$$

The soil pH was calculated using the following formulas<sup>24</sup>.

$$\text{pH} = 7.15 + 1.59*\text{CBT}' \quad (n = 221, R^2 = 0.85, \text{RMSE} = 0.52, P < 0.0001) \quad (6)$$

$$\text{CBT}' = -\log \frac{Ic + IIa' + IIb' + IIc' + IIIa' + IIIb' + IIIc'}{Ia + Ib + Ic} \quad (7)$$

SWC is well correlated with  $\text{MBT}'$  when  $\text{IR}_{6\text{ME}} > 0.5$ , and these proxies were calculated using the following expressions:

$$\text{MBT}' = \frac{(Ia + Ib + Ic)}{(Ia + Ib + Ic + IIa + IIa' + IIb + IIb' + IIc + IIc' + IIIa + IIIa')} \quad (8)$$

$$\text{IR}_{6\text{ME}} = \frac{\sum(\text{C6} - \text{methylated brGDGTs})}{\sum \text{brGDGTs}} \quad (9)$$

$$\text{MBT}'_{6\text{ME}} = \frac{(Ia + Ib + Ic)}{(Ia + Ib + Ic + IIa' + IIb' + IIc' + IIIa')} \quad (10)$$

where the Roman numerals indicate different brGDGTs structures (Supplementary Fig. 1).

### Principal component analysis (PCA)

CANOCO version 5 software was utilized to reveal the relationships among various environmental factors. The first PCA figure (Fig. 3a) was

generated for the environmental factors MAT, MAPc, SWC, and pH. These variables are based on the same dataset (238 LSPs samples from Weinan profile) without any data transformation. The second PCA figure (Fig. 3b) was generated for the environmental factors MAT, MAP (based on <sup>10</sup>Be), SWC and pH, which were all in the transition of the glacial–interglacial after 430 ka BP on the CLP. As the two LSPs profile had similar sedimentation rates, we obtained the same chronological control through linear interpolation in those transition periods. All datasets passed the Gaussian distribution test in this study.

### Cross wavelet analysis

Compared with wavelet special analysis, cross wavelet analysis is even more complicated. The wavelet cross-spectrum can be defined as follows:

$$CS(b, a) = m_{1c}(b, a)m_{2c}(b, a) \tag{11}$$

where  $m_{1c}$  and  $m_{2c}$  describe the covarying fractions of the overall spectra given by:

$$m_1(b, a) = m_{1c}(b, a) + m_{1i}(b, a) \tag{12}$$

$$m_2(b, a) = m_{2c}(b, a) + m_{2i}(b, a) \tag{13}$$

where  $m_{1i}$  and  $m_{2i}$  are independent contributions to the variance.

Overall, this is a multipart function that may be decomposed into amplitude and phase:

$$CS(b, a) = |CS(b, a)|\exp(i \arg(CS(b, a))) \tag{14}$$

In this study, a and b represent the array of reconstructed MAPc and the Sanbao speleothem  $\delta^{18}O$ , respectively.

### Multiple regression linear model

To compare the precision of the DLNN-MAP model we established, we set up a multiple regression linear model based on all 6-methyl brGDGTs except Ib. The basis of the model is defined as:

$$y = a + b_1x_1 + b_2x_2 \dots + b_nx_n \tag{15}$$

where  $y$  represents MAP,  $x$  represents all 6-methyl brGDGTs and Ia and Ic, and  $a, b_1, b_2 \dots b_n$  represent the partial regression coefficients.  $n$  represents the number of 6-methyl we entered into the model (in this study,  $n = 8$ ).

The multiple correlation coefficient ( $R$ ) was defined as follows:

$$R = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \tag{16}$$

where  $y_i$  represents the actual observed value,  $\hat{y}_i$  represents the calculation value and  $\bar{y}$  represents the mean value of all observed data.

The  $t$  statistic is used to test the validity of regression coefficients, and it can be defined as follows:

$$t_{b_j} = \frac{b_j}{S_{b_j}} \tag{17}$$

$$S_{b_j} = \sqrt{P_{jj}^*s} \tag{18}$$

**Table 2 | List of the parameters of the multiple linear model**

Model	Coefficient	$t$	Significance	VIF
a	1428.527	41.153	0.000	
b <sub>1</sub>	0.116	0.526	0.599	2.561
b <sub>2</sub>	1524.901	2.641	0.009	2.638
b <sub>3</sub>	-2070.924	-7.983	0.000	2.792
b <sub>4</sub>	-941.825	-1.190	0.235	2.365
b <sub>5</sub>	4507.380	0.432	0.666	2.191
b <sub>6</sub>	-1158.136	-3.672	0.000	2.347
b <sub>7</sub>	-3786.251	-5.061	0.000	1.061
b <sub>8</sub>	11643.395	-0.421	0.674	1.305

$$s = \sqrt{\frac{1}{n - m - 1} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{19}$$

$$P = (p_{jj})^{-1} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \tag{20}$$

where  $b_j$  represents the regression coefficient of  $x_j$ ,  $n$  represents the number of samples and  $m$  represents the number of variables.

The  $F$  statistic is used to test the linear relationship of variables and can be defined as follows:

$$F = \frac{1}{ms^2} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \tag{21}$$

The variance inflation factor (VIF) is used to measure collinearity between variables and can be defined as follows:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{22}$$

As shown in Supplementary Fig. 5, we found no obvious collinearity between different variables. However, there are fewer contributions in IIc', IIIa', IIIb', and IIIc' in the multiple regression linear model we established, and the value of  $t$  cannot attain the 95% confidence level (Table 2). The results of both the training dataset and extrapolated experimental dataset (Supplementary Fig. 6), although they seem good ( $R^2 = 0.44$  and  $0.46$ , respectively), still have a considerable gap compared with the DLNN-MAP model. Especially when MAP > 1500 mm, the multiple linear model is unable to forecast the real MAP. These results all indicate that the MAP influence on the brGDGTs compounds is not a simple linear relationship; instead, we suggest that there are complex pilot processes between them.

### DLNN models

DLNNs usually contain an input layer, a few hidden layers, and an output layer. A conceptual structure of the DLNN model was established for forecasting MAP values. The first layer accepts input signals that are various combinations of brGDGTs. The relationships among different variables are processed and analyzed in the hidden layers. The final class output is presented in the output layer; in this study, the output is the MAP reconstruction at the study site.

The rectified linear unit (ReLU) activation function is applied in each neuron of the hidden layer, which is computationally simpler than the traditionally applied sigmoid function. The function of the ReLU

activation function is given as follows:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} = \max(0, x) \quad (23)$$

where  $x$  represents an input signal to a neuron and  $f$  represents the activation function.

Furthermore, the bias between the measured and forecasted output values is reflected by the loss function. The loss function applied herein is the MAE (mean absolute error), which is given as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^n |T - Y| \quad (24)$$

where  $N$  is the number of training data points, and  $T$  and  $Y$  represent the measured output value and the forecasted class value, respectively.

To realize the backpropagation framework, the derivative of the ReLU activation function needs to be acquired. According to the definition of the ReLU, the derivative is shown as follows.

$$f'(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (25)$$

Given a minibatch of  $m$  training samples obtained from the training set  $\{x(1), x(2), \dots, x(m)\}$  and their corresponding targets  $T(i)$  ( $i = 1, 2, \dots, m$ ), the gradient used in the backpropagation framework is shown as follows:

$$f = \frac{1}{m} \sum_{i=1}^n \frac{\partial L}{\partial w} \quad (26)$$

where  $L$  is the loss function;  $w$  represents the network weights; and  $n = 1$  is the number of output values (MAP).

In addition, considering that the adaptive moment estimation algorithm (Adam) was proven to be an effective neural network training method with a fast convergence speed and great classification performance, we applied this algorithm to train the DLNN model for MAP forecasting in this study. Adam has two biased equations, which are shown as follows:

$$a = \rho_1 a + (1 - \rho_1)g \quad (27)$$

$$b = \rho_2 b + (1 - \rho_2)g \times g \quad (28)$$

where  $\rho_1 = 0.9$  and  $\rho_2 = 0.999$  are exponential decay rates;  $g$  is the gradient; and  $\times$  represents an elementwise product operator.

After this calculation, the correct biases in the above two moments are given as follows:

$$a_c = \frac{a}{1 - \rho_1^t} \quad (29)$$

$$b_c = \frac{b}{1 - \rho_2^t} \quad (30)$$

where  $t$  represents the current time step.

Moreover, the update of the network weights is shown as follows:

$$\Delta_w = -\lambda \frac{a_c}{\sqrt{b_c + \epsilon}} \quad (31)$$

where  $\lambda = 0.001$  represents the learning rates and  $\epsilon = 10^{-8}$  is a constant used to ensure numerical stability.

Eventually, the DLNN parameters can be updated according to the following formula.

$$w = w + \Delta_w \quad (32)$$

### brGDGT-MAP models

We entered 9 brGDGTs compounds (all 6-methyl brGDGTs; each compound entered in the model is the percentage of all brGDGTs in the surface soil) into the input layer of the DLNN; these compounds are closely related to soil moisture. Then, we selected 533 surface soil samples as the training dataset and 179 surface soil samples as the validation dataset, both of which satisfied the principle of randomness. We assessed the precision of the model using forecast data  $R^2$  and root mean square error (RMSE) values.

Through several parameters applied in the DLNN model, we found that the frequency of training and the number of neurons play the most significant roles in the brGDGT-MAP models. In addition, four hidden layers containing the other DLNN parameters allow the model to become more stable (detailed parameters are shown in Supplementary Fig. 7). To test the best frequency of training and the number of neurons in each hidden layer, we set a series of gradients to test the model to find the most suitable combination. As shown in Supplementary Fig. 8, for the frequency of training, we set the minimum and maximum training times to 1000 and 1500, respectively, with 100 times as the interval. We also set the numbers of neurons from 160 to 260 with a 20-neuron interval.

Testing the weights of different compounds in the DLNN model and determining whether it was essential to eliminate some compounds that may make the dataset redundant were also required. Based on the model in which the lb parameter was removed, we also set a series of experiments to test the effects of the different 6-methyl isomers on the predicted MAP values. Then, we made seven attempts to test the forecast effect of the brGDGT-MAP models by removing the lc, Ila', I Ib', Ilc', IIIa', IIIb', and IIIc' parameters (Supplementary Fig. 9). Then, we obtained the best brGDGT-MAP model (Supplementary Fig. 10).

### Comparison of various ANN structures

To improve the accuracy of our brGDGT-MAP models and the models' universality, we also tested more complex ANN structures and then compared them with our DLNN models.

**RNN.** A recurrent neuron network (RNN) is an artificial neural network in which nodes are directionally connected into loops. The essential feature of RNN is that there are both internal feedback connections and feedforward connections between processing units. The inner structure of RNN is similar to that of the human brain, which can learn to transform a lifetime of sensory input streams into an efficient sequence of motor outputs (Supplementary Fig. 11a). Therefore, the basis of the RNN is defined as follows:

$$h_t = f(U * X_t + W * h_{t-1}) \quad (33)$$

$$o_t = \text{softmax}(V h_t) \quad (34)$$

where  $X_t$  represents the input value at time  $t$ ;  $o_t$  represents the output value at time  $t$ ;  $h_t$  represents the memory value at time  $t$ ; and  $U$ ,  $V$ , and  $W$  are the parameters of this network. For the motivative function, we chose softmax.

**LSTM.** Long short-term memory networks (LSTM) are a special type of RNN that can learn long-term dependence and contain three gates (forget gate, input gate and output gate) and one memory cell. The

horizontal line above the box is called the cell state, and it acts as a conveyor belt to control the flow of information to the next moment (Supplementary Fig. 11b). Therefore, the basis of LSTM is defined as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (35)$$

where  $C_{t-1}$  represents the knowledge state of the model at time  $t-1$  and  $\tilde{C}_t$  represents the newly acquired information after entering new observations.  $f_t$  and  $i_t$  represent the weight parameters of  $C_{t-1}$  and  $\tilde{C}_t$ , respectively.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (36)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (37)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (38)$$

where  $h_{t-1}$  represents the output value at time  $t-1$  and  $x_t$  represents the new input value at time  $t$ .  $W_f$  represents the motivative function in this study. We used  $\tanh$  as the motivative function when our model was learning. Each new input may not have a positive impact on the machine, but it may also have a negative impact.  $b_f$ ,  $b_i$  and  $b_c$  represent the random disturbances (white noise).

**GRU.** As mentioned above, the LSTM is proposed to overcome RNN's inability to address remote dependence and the gate recurrent unit (GRU), a variant of the LSTM, keeps the effect of the LSTM while making the structure simpler.

Compared with the LSTM, the GRU only has two gates (update ( $z_t$ ) and reset ( $r_t$ ) gates). The update gate is used to control the degree to which the state information at the previous moment is brought into the current state. The larger the value of the update gate is, the more state information at the previous moment is brought in. The reset gate is used to control the degree to which the state information at the previous moment is ignored (Supplementary Fig. 11c). Therefore, the basis of the LSTM is defined as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (39)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (40)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t]) \quad (41)$$

$$h_t = (1 - z_t) * r_t * h_{t-1} + z_t * \tilde{h}_t \quad (42)$$

$$y_t = \sigma(W_o \cdot h_t) \quad (43)$$

where  $[\ ]$  represents the connection of two vectors and  $*$  represents the multiplication of matrix elements. The  $x_t$  and  $y_t$  represent the input and output values at time  $t$ , respectively.

It can be seen from the above formula that the parameters to be learned are the weight parameters of  $W_r$ ,  $W_z$ ,  $W_h$ , and  $W_o$ . The first three weights are spliced; therefore, they need to be segmented during learning. These can be defined as follows:

$$W_r = W_{rx} + W_{rh} \quad (44)$$

$$W_z = W_{zx} + W_{zh} \quad (45)$$

$$W_h = W_{hx} + W_{hh} \quad (46)$$

As we can find in the RNN, LSTM, and GRU models we reconstructed (Supplementary Fig. 12), the training datasets all show extraordinarily high  $R^2$  values (0.99, 0.99, and 0.99, respectively) and low RMSE values (0.36, 0.23, and 0.16, respectively). However, the validation datasets do not show good prediction ability compared with the DLNN. These results indicate that the two ANN structures are not suitable for MAP prediction based on brGDGTs, although their inner structures are more complex than those of the DLNN. The reason we suggested is that the RNN, LSTM and GRU are more appropriate to the massive amounts of data and the data that have obvious spatio-temporal characteristics. The great prediction precision in the training dataset and the poor performance in the extrapolated datasets indicate that the models based on the RNN, LSTM and GRU have significant overfitting. As a result, compared with other ANN structures, we concluded that our DLNN model is the most suitable one to forecast MAP based on brGDGTs.

### Environmental indicators of *n*-alkanes proxies

Long-chain *n*-alkanes in plant leaf waxes are universal in terrestrial environments and can deliver signals of variations in plant sources and past climate. They are widely distributed in surface soils and Quaternary sediments, especially in LPSS. In this study, due to the insufficient samples in Weinan profile, we only analyzed *n*-alkanes components for 40 LPS samples, which contain ages between 340 and 430 ka BP.

### Instrumental measurements

For the apolar fractions, a total of 40 samples in this study, mainly containing *n*-alkanes, were all investigated utilizing a Shimadzu 2010 gas chromatograph (GC) equipped with a flame ionization detector (FID) and a DB-5 fused silica capillary column (60 m  $\times$  0.32 mm  $\times$  0.25  $\mu$ m film thickness) with helium as the carrier gas. The temperature of the GC oven was enhanced from 70 to 300  $^{\circ}$ C at a rate of 3  $^{\circ}$ C/min. Then, this temperature (300  $^{\circ}$ C) was maintained for 30 min. Finally, the concentrations of the *n*-alkane homologs were evaluated by assessing the peak area of the *n*-alkanes to that of the internal standard (cholane).

### Long-term paleoclimatic change

The carbon preference index (CPI) evaluates the relative abundances of odd vs. even-numbered *n*-alkanes. The CPI increases as the environmental aridity increases. The CPI indicated warm-wet periods and cold-dry periods in paleoclimate and corresponded well with the loess-paleosol cycle<sup>52</sup>. The average chain length (ACL) value is the weighted average of the different carbon chain lengths. The lower ACL value corresponds to the lower temperature in the research of LPSS. The variations in the ACL value have good coordination with the magnetic susceptibility and particle size. The *n*-alkane CPI<sup>53</sup> and ACL<sup>54</sup> are calculated as follows:

$$CPI(1) = \frac{(C_{23} + C_{25} + C_{27} + C_{29} + C_{31}) + (C_{25} + C_{27} + C_{29} + C_{31} + C_{33})}{2(C_{24} + C_{26} + C_{28} + C_{30} + C_{32})} \quad (47)$$

$$CPI(2) = \frac{1}{2} \left( \frac{C_{25} + C_{27} + C_{29} + C_{31} + C_{33}}{C_{24} + C_{26} + C_{28} + C_{30} + C_{32}} + \frac{C_{25} + C_{27} + C_{29} + C_{31} + C_{33}}{C_{26} + C_{28} + C_{30} + C_{32} + C_{34}} \right) \quad (48)$$

$$ACL = \frac{23C_{23} + 25C_{25} + 27C_{27} + 29C_{29} + 31C_{31} + 33C_{33}}{C_{23} + C_{25} + C_{27} + C_{29} + C_{31} + C_{33}} \quad (49)$$

Figure 13 shows the variations in CPI (Supplementary Fig. 13a) and ACL (Supplementary Fig. 13b) values in the Weinan profile from 340 to 430 ka BP. Compared with the MAP (Supplementary Fig. 13c) and SWC (Fig. 2e) reconstructions based on brGDGTs, we found that they all had a peak at ~350 ka BP, which indicates relatively high soil moisture at approximately 350 ka BP.

### MAP reconstruction in the XRD section

In this section, we test the brGDGT-MAP model in the Xiangride (XRD) profile, which is located in the margin region of the East Asian monsoon (Fig. 1). With robust chronological control, we reconstructed the rainfall changes in 7000 years BP (Supplementary Fig. 14b). We found that MAP was ~200 mm in the late Holocene, which approaches multiple modern observations in this region (180 mm). Moreover, we suggest that this region experienced the most humid period in the mid-Holocene, when the rainfall reached 600 mm. Afterward, the precipitation declined from 6000 to 4000 years BP and then increased and reached a peak value at ~3000 years BP. Then, it had a drought trend until modern times.

We discovered that our brGDGT-MAP model could precisely capture rainfall dynamics based on the Weinan profile (Supplementary Fig. 14a) and XRD profile (Supplementary Fig. 14b). Combined with the most acceptable rainfall records in the Holocene (i.e.,  $^{10}\text{Be}$  (Supplementary Fig. 14c), pollen in Gonghai (Fig. 1 and Supplementary Fig. 14d), and Dongge cave  $\delta^{18}\text{O}$  (Fig. 1 and Supplementary Fig. 14e)), we found the same precipitation peak values in the early Holocene and mid-Holocene. In addition, they all revealed a drought trend throughout the whole Holocene. We suggest that brGDGTs can become a robust proxy to reconstruct precipitation in the Holocene.

### Data availability

The relevant data of Weinan LPSs profile are provided in the Supplementary Information Data file. Source data are provided with this paper.

### Code availability

The code used in this study is available from the corresponding author upon request.

### References

- Heller, F. & Tunngsheng, L. Magnetism of Chinese loess deposits. *Geophys. J. Int.* **77**, 125–141 (1984).
- Li, G. et al. Paleoclimatic changes and modulation of East Asian summer monsoon by high-latitude forcing over the last 130,000 years as revealed by independently dated loess-paleosol sequences on the NE Tibetan Plateau. *Quat. Sci. Rev.* **237**, 106283 (2020).
- An, Z. The history and variability of the East Asian paleomonsoon climate. *Quat. Sci. Rev.* **19**, 171–187 (2000).
- Wang, Y. J. et al. A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345 (2001).
- Kukla, G. Loess stratigraphy in central China. *Quat. Sci. Rev.* **6**, 191–219 (1987).
- Guo, Z. et al. Summer monsoon variations over the last 1.2 Ma from the weathering of loess-soil sequences in China. *Geophys. Res. Lett.* **27**, 1751–1754 (2000).
- Liu, T., Ding, Z. & Rutter, N. Comparison of Milankovitch periods between continental loess and deep sea records over the last 2.5 Ma. *Quat. Sci. Rev.* **18**, 1205–1212 (1999).
- Beck, J. W. et al. A 550,000-year record of East Asian monsoon rainfall from  $^{10}\text{Be}$  in loess. *Science* **360**, 877 (2018).
- Sun, Y. et al. Diverse manifestations of the mid-Pleistocene climate transition. *Nat. Commun.* **10**, 352 (2019).
- Li, T. et al. Continued obliquity pacing of East Asian summer precipitation after the mid-Pleistocene transition. *Earth Planet. Sci. Lett.* **457**, 181–190 (2017).
- Clemens, S. C. et al. Precession-band variance missing from East Asian monsoon runoff. *Nat. Commun.* **9**, 3364 (2018).
- Walker, M. *Quaternary Dating Methods* (J. Wiley, 2005).
- Cheng, H. et al. The Asian monsoon over the past 640,000 years and ice age terminations. *Nature* **534**, 640 (2016).
- Wang, Y. et al. Millennial- and orbital-scale changes in the East Asian monsoon over the past 224,000 years. *Nature* **451**, 1090–1093 (2008).
- Lachniet, M. S. Illuminating the meaning of Asian monsoon cave speleothem records. *Paleoceanogr. Paleoclimatology* **35**, e2019P–e3841P (2020).
- Yang, H., Johnson, K. R., Griffiths, M. L. & Yoshimura, K. Interannual controls on oxygen isotope variability in Asian monsoon precipitation and implications for paleoclimate reconstructions. *J. Geophys. Res.: Atmospheres* **121**, 8410–8428 (2016).
- Lu, H. Y. et al. Phytoliths as quantitative indicators for the reconstruction of past environmental conditions in China I: phytolith-based transfer functions. *Quat. Sci. Rev.* **25**, 945–959 (2006).
- Lu, H. Y., Wu, N. Q., Liu, K. B., Jiang, H. & Liu, T. S. Phytoliths as quantitative indicators for the reconstruction of past environmental conditions in China II: palaeoenvironmental reconstruction in the Loess Plateau. *Quat. Sci. Rev.* **26**, 759–772 (2007).
- Youfeng, N., Weiguo, L. & Zhisheng, A. A 130-ka reconstruction of precipitation on the Chinese Loess Plateau from organic carbon isotopes. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **270**, 59–63 (2008).
- Sun, J., Lü, T., Zhang, Z., Wang, X. & Liu, W. Stepwise expansions of C4 biomass and enhanced seasonal precipitation and regional aridity during the Quaternary on the southern Chinese Loess Plateau. *Quat. Sci. Rev.* **34**, 57–65 (2012).
- Weijers, J. W. H. et al. Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environ. Microbiol.* **8**, 648–657 (2006).
- Weijers, J. W. H., Steinmann, P., Hopmans, E. C., Schouten, S. & Sinninghe Damsté, J. S. Bacterial tetraether membrane lipids in peat and coal: testing the MBT–CBT temperature proxy for climate reconstruction. *Org. Geochem.* **42**, 477–486 (2011).
- Naafs, B. D. A., Gallego-Sala, A. V., Inglis, G. N. & Pancost, R. D. Refining the global branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration. *Org. Geochem.* **106**, 48–56 (2017).
- De Jonge, C. et al. Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction. *Geochim. Cosmochim. Ac.* **141**, 97–112 (2014).
- Weijers, J. W., Schouten, S., van den Donker, J. C., Hopmans, E. C. & Damsté, J. S. S. Environmental controls on bacterial tetraether membrane lipid distribution in soils. *Geochim. Cosmochim. Ac.* **71**, 703–713 (2007).
- Peterse, F. et al. Revised calibration of the MBT–CBT paleo-temperature proxy based on branched tetraether membrane lipids in surface soil. *Geochim. Cosmochim. Ac.* **96**, 215–229 (2012).
- Lu, H. et al. 800-kyr land temperature variations modulated by vegetation changes on Chinese Loess Plateau. *Nat. Commun.* **10**, 1–10 (2019).
- Tierney, J. E. et al. Northern hemisphere controls on tropical southeast African Climate during the past 60,000 years. *Science* **322**, 252 (2008).
- Weijers, J. W. H., Schefuß, E., Schouten, S. & Damsté, J. S. S. Coupled thermal and hydrological evolution of tropical Africa over the last deglaciation. *Science* **315**, 1701 (2007).
- Halamka, T. A. et al. Oxygen limitation can trigger the production of branched GDGTs in culture. *Geochemical Perspect. Lett.* **19**, 36–39 (2021).

31. Weber, Y. et al. Redox-dependent niche differentiation provides evidence for multiple bacterial sources of glycerol tetraether lipids in lakes. *Proc. Natl Acad. Sci. USA* **115**, 10926 (2018).
32. Dearing Crampton-Flood, E., Tierney, J. E., Peterse, F., Kirkels, F. M. S. A. & Sinninghe Damsté, J. S. BayMBT: a Bayesian calibration model for branched glycerol dialkyl glycerol tetraethers in soils and peats. *Geochim. Cosmochim. Ac.* **268**, 142–159 (2020).
33. Martínez-Sosa, P. et al. A global Bayesian temperature calibration for lacustrine brGDGTs. *Geochim. Cosmochim. Ac.* **305**, 87–105 (2021).
34. Raberg, J. H. et al. Revised fractional abundances and warm-season temperatures substantially improve brGDGT calibrations in lake sediments. *Biogeosciences* **18**, 3579–3603 (2021).
35. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
36. Dang, X., Yang, H., Naafs, B. D. A., Pancost, R. D. & Xie, S. Evidence of moisture control on the methylation of branched glycerol dialkyl glycerol tetraethers in semi-arid and arid soils. *Geochim. Cosmochim. Ac.* **189**, 24–36 (2016).
37. Slessarev, E. W. et al. Water balance creates a threshold in soil pH at the global scale. *Nature* **540**, 567–569 (2016).
38. Guo, J. et al. Soil pH and aridity influence distributions of branched tetraether lipids in grassland soils along an aridity transect. *Org. Geochem.* **164**, 104347 (2021).
39. Kukla, G. et al. Pleistocene climates in China dated by magnetic susceptibility. *Geology* **16**, 811–814 (1988).
40. Ding, Z. L. et al. Stacked 2.6-Ma grain size record from the Chinese loess based on five sections and correlation with the deep-sea  $\delta^{18}\text{O}$  record. *Paleoceanography* **17**, 1–5 (2002).
41. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, 1003 (2005).
42. Laskar, J. et al. A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
43. Liu, J. et al. Holocene East Asian summer monsoon records in northern China and their inconsistency with Chinese stalagmite  $\delta^{18}\text{O}$  records. *Earth-Sci. Rev.* **148**, 194–208 (2015).
44. Tang, C., Yang, H., Dang, X. & Xie, S. Comparison of paleo-temperature reconstructions using microbial tetraether thermometers of the Chinese loess-paleosol sequence for the past 350000 years. *Sci. China Earth Sci.* **60**, 1159–1170 (2017).
45. Li, G. et al. Differential ice volume and orbital modulation of quaternary moisture patterns between Central and East Asia. *Earth Planet. Sc. Lett.* **530**, 115901 (2020).
46. Liang, J. et al. Vegetation effects on temperature calibrations of branched glycerol dialkyl glycerol tetraether (brGDGTs) in soils. *Org. Geochem.* **127**, 1–11 (2019).
47. Murakami, T. et al. Centennial- to millennial-scale climate shifts in continental interior Asia repeated between warm–dry and cool–wet conditions during the last three interglacial states: evidence from uranium and biogenic silica in the sediment of Lake Baikal, south-east Siberia. *Quat. Sci. Rev.* **52**, 49–59 (2012).
48. Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **5**, 170191 (2018).
49. Tang, C. et al. Tropical and high latitude forcing of enhanced megadroughts in Northern China during the last four terminations. *Earth Planet. Sc. Lett.* **479**, 98–107 (2017).
50. Hopmans, E. C. et al. A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. *Earth Planet. Sc. Lett.* **224**, 107–116 (2004).
51. Yang, H. et al. Correlations between microbial tetraether lipids and environmental variables in Chinese soils: optimizing the paleo-reconstructions in semi-arid and arid regions. *Geochim. Cosmochim. Ac.* **126**, 49–69 (2014).
52. Zhang, Z., Zhao, M., Eglinton, G., Lu, H. & Huang, C. Leaf wax lipids as paleovegetational and paleoenvironmental proxies for the Chinese Loess Plateau over the last 170kyr. *Quat. Sci. Rev.* **25**, 575–594 (2006).
53. Marzi, R., Torkelson, B. E. & Olson, R. K. A revised carbon preference index. *Org. Geochem.* **20**, 1303–1306 (1993).
54. Badewien, T., Vogts, A. & Rullkötter, J. n-Alkane distribution and carbon stable isotope composition in leaf waxes of C3 and C4 plants from Angola. *Org. Geochem.* **89–90**, 71–79 (2015).
55. Sun, W., Zhao, S., Pei, H. & Yang, H. The coupled evolution of mid- to late Holocene temperature and moisture in the southeast Qaidam Basin. *Chem. Geol.* **528**, 119282 (2019).
56. Chen, F. et al. East Asian summer monsoon precipitation variability since the last deglaciation. *Sci. Rep.* **5**, 11186 (2015).
57. Spratt, R. M. & Lisiecki, L. E. A Late Pleistocene sea level stack. *Clim.* **12**, 1079–1092 (2016).
58. Xiao, W. et al. Global calibration of a novel, branched GDGT-based soil pH proxy. *Org. Geochem.* **89–90**, 56–60 (2015).
59. Yang, H. et al. The 6-methyl branched tetraethers significantly affect the performance of the methylation index (MBT') in soils from an altitudinal transect at Mount Shennongjia. *Org. Geochem.* **82**, 42–53 (2015).
60. Lei, Y., Yang, H., Dang, X., Zhao, S. & Xie, S. Absence of a significant bias towards summer temperature in branched tetraether-based paleothermometer at two soil sites with contrasting temperature seasonality. *Org. Geochem.* **94**, 83–94 (2016).
61. Wang, H., An, Z., Lu, H., Zhao, Z. & Liu, W. Calibrating bacterial tetraether distributions towards in situ soil temperature and application to a loess-paleosol sequence. *Quat. Sci. Rev.* **231**, 106172 (2020).
62. Wang, H., Liu, W. & Lu, H. Appraisal of branched glycerol dialkyl glycerol tetraether-based indices for North China. *Org. Geochem.* **98**, 118–130 (2016).
63. Ding, S. et al. Distribution of branched glycerol dialkyl glycerol tetraethers in surface soils of the Qinghai-Tibetan Plateau: implications of brGDGTs-based proxies in cold and dry regions. *Biogeosciences* **12**, 3141–3151 (2015).
64. Wang, M. et al. Utility of brGDGTs as temperature and precipitation proxies in subtropical China. *Sci. Rep.* **8**, 194 (2018).
65. Wang, M., Zheng, Z., Zong, Y., Man, M. & Tian, L. Distributions of soil branched glycerol dialkyl glycerol tetraethers from different climate regions of China. *Sci. Rep.* **9**, 2761 (2019).

## Acknowledgements

We thank the three anonymous reviewers and the editor for their extensive comments, which have no doubt helped to improve the quality of this manuscript. We thank Changyan Tang and Guoqiao Xiao for field sampling. This study is supported by the National Natural Science Foundation of China (Nos. 41790422 and 41830319). We sincerely thank the Institute for Planets Company and its designers, Borong Zheng and Rui Du, for their help on image design. The learning experiments were run using AlforScience produced by Beijing Diji Tech, a platform that accelerates the implementation and explanation of deep-learning algorithms.

## Author contributions

Y.Z.: study conceived and developed, date analysis, writing the 1st draft. H.L.: conceptualization, funding acquisition, supervision, writing, review, and editing. H.Y.: conceptualization, writing, review, and editing. H.W.: writing, review, and editing. W.Z., Z.Z., M.H., and W.L.: resources.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-33105-2>.

**Correspondence** and requests for materials should be addressed to Hongyan Liu or Huan Yang.

**Peer review information** *Nature Communications* thanks Sarah Coffinet, Guoqiang Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022