

Supporting Information

Ali et al. 10.1073/pnas.1203467109

SI Text

Main Characteristics of the Lacustrine Deposits. The nine lakes are located along a 350-km east–west transect southeast of the James Bay area (50°N, 80°W; Fig. S1). The lakes were chosen because of their similar size, maximum depth, and shape (Table S1).

Chronologies and Sedimentation Rates. Accelerated mass spectrometry (AMS) radiocarbon measurements were used to establish the chronology of the sites. The measurements were applied to terrestrial plant remains (i.e., needles, tree seeds) when available. ¹⁴C measurements were then calibrated in years before present (hereafter cal yBP), using the CALIB program version 6.0 (1) and based on the intcal09 dataset (2).

The age–depth models were computed using the MCAgeDepth program (3) (www.uidaho.edu/cnr/paleoecologylab/software), which applies a Monte Carlo resampling technique to assess median ages and to generate confidence intervals (CI) around the fit on the basis of the probability distribution of each date given by the CALIB program (Fig. S2). The resulting spline-based age–depth models, associated with the CI from the Monte Carlo simulations, are more robust than those constructed on the basis of ages inferred from the median values of calibrated dates or by the line intersect method.

Reconstruction of Regional Biomass Burning. Individual charcoal accumulation rate (CHAR) series were homogenized (4) to reduce the influence of sedimentation rate and potential taphonomic biases linked to the sequestration of charcoal in the lacustrine deposits. Homogenized series were then pooled to build the regional biomass burning (*RegBB*). The procedure followed three different steps: (i) rescaling initial CHAR values using a min–max transformation (Eq. S1), (ii) homogenizing the variance using the Box–Cox transformation (Eq. S2), and (iii) rescaling the values to Z-scores (Eq. S3).

Min–max transformation is computed as

$$CHAR'_i = (CHAR_i - CHAR_{\min}) / (CHAR_{\max} - CHAR_{\min}), \quad [S1]$$

where $CHAR'_i$ was the min–max-transformed value of the i th sample of a particular record. $CHAR_{\min}$ and $CHAR_{\max}$ corresponded to the maximum and minimum values of the $CHAR_i$ series.

Box–Cox transformation is computed as

$$CHAR''_i = \begin{cases} ((CHAR'_i + \alpha)^\lambda - 1) / \lambda & \text{if } \lambda \neq 1 \\ \log(CHAR'_i + \alpha) & \text{if } \lambda = 1, \end{cases} \quad [S2]$$

where $CHAR''_i$ were the transformed series, α was a small positive constant that avoids singularity when $CHAR'_i = 0$ and $\lambda = 0$ (here taken equal to 0.01), and the transformation parameter λ was selected so that the $CHAR''_i$ frequency distribution was approximately normal. The λ -value was estimated by maximum likelihood following Venables and Ripley (5).

Z-score transformation was done as follows:

$$CHAR'''_i = (CHAR''_i - \overline{CHAR''}_{i(7ky)}) / \sigma \overline{CHAR''}_{i(7ky)}. \quad [S3]$$

Eq. S3 enables all transformed $CHAR'''_i$ series to have a common mean and variance. $\overline{CHAR''}_{i(7ky)}$ was the mean of $CHAR''_i$ series

over the last 7,000 y and $\sigma \overline{CHAR''}_{i(7ky)}$ the SD over the same period.

In the case of a pooled record of multiple lakes, the values of *RegBB* were adjusted to the changing number of lake samples through time, using

$$RegBB_t = \sum_i^n CHAR'''_i / n_t, \quad [S4]$$

where n_t was the number of sampled cores at time t . Smoothing was performed using locally weighted regression or LOWESS, allowing us to minimize the influence of outlier values. We assessed the significance of changes in *RegBB* with the help of bootstrap confidence intervals (BCI) (90%). Detected trends in *RegBB* were confirmed by box-plot analysis carried out at each 1,000 y. The significance of differences between *RegBB* between millennia was assessed using the Wilcoxon rank sum test.

An overall relationship existed between the sedimentation rate and charcoal concentration and influx (Fig. S4). Maximum sedimentation rate occurred between 5 and 3 kyBP. This change in sedimentation rate could be related to the increase in superficial transport and erosion processes triggered by the significant increase in precipitation during this period, notably in spring during the early fire season. However, between 5.5 and 1.5 kyBP, no significant change in long-term biomass burning occurred (Fig. 1A, main text). This suggests a long-term independence between the sedimentation rate and the charcoal concentration and influx fluctuations.

Correlation Between *RegBB* and a Stand-Replacing Fire History. This section deals with a correlation analysis between the annual proportion of area burned (i.e., burn rate) in the study area, inferred from a stand-replacing fire history reconstruction, and *RegBB*. Data on the burn rates (i.e., the inverse of the fire cycle) are those of Bergeron et al. (6), their figure 6. Below is a summary of the procedure for computation of the burn rates described by Bergeron et al. (6).

Briefly, a time-since-fire (TSF) distribution, from the transition zone between the Balsam fir–white birch and Spruce–feather moss bioclimatic domains, was obtained from a postfire stand initiation map created using field and archival data digitalized and included in a geographic information system database. The area covered by this map encompasses a territory of 15,000 km² (78°30'W to 79°30'W and 48°00'N to 50°00'N). Fires that occurred after 1880 were precisely dated and mapped. For older fire years, an estimate of the area burned was obtained by dendrochronological dating and interpolation. Because of the lack of precision in estimating the fire date from the field data, all fires were binned in 10-y age classes. Survival analysis was then used to estimate the fire cycle for different time periods (7) (Proc LIFEREG; SAS Institute 1990). In this analysis, TSF estimates were considered censored when no accurate date could be attributed to the original fire; in these cases, a minimum TSF estimate was used, i.e., the age of the oldest individual when no tree cohort could be clearly identified. Under a negative exponential distribution ($S(t) = \exp^{-(t/b)}$), the value of the “survival parameter” b is the mean of the distribution (7) and corresponds to the fire cycle of the study area. The survival analysis was repeated “sequentially,” by decade, starting with the present and going back in time by eliminating processed decades. The entire data distribution (from 1999 backward) was thus used to evaluate the fire cycle of the study area in 1999. Then, areas that

burned during the decade of 1999–1990 were eliminated and the fire cycle was recomputed; eliminated areas were considered as censured as there is no information on when they had last burned before 1990. The same procedure was repeated, eliminating two, three, and four decades and so on, until all of the decades back to 1770 were eliminated. Low site replication before 1770 prevented us from going farther back in time. Once fire cycle estimates were determined, they were converted into burn rate estimates, using the inverse of the fire cycles. This procedure allowed us to evaluate past variations in the burn rates with a resolution of 10 y (Fig. S3A).

Past variations in the burn rates were correlated to *RegBB* over the period of overlap. Owing to uncertainties in calibration dating, the *RegBB* reconstruction was shifted backward and forward to optimize the correlation with the burn rate estimates (8). The optimal correlation found was obtained by attributing the date of 1980 to the end of the *RegBB* reconstruction (which is within the 30-y range suggested by ref. 8). Visual inspection suggested that burn rate estimates and *RegBB* had a significant amount of variance in common. Low burn rate estimates in the early 1800s, late 1800 to early 1900s, and late 20th century corresponded relatively well to troughs in the *RegBB* reconstruction, and vice versa. The Spearman rank correlation between the two series is 0.73, with a 95% bootstrap confidence interval [0.30, 0.89] ($n = 22$ decades, taking into account autocorrelation in data) (9). The functional form between burn rate estimates and *RegBB* is close to being linear (Fig. S3B).

Reconstruction of Fire Events and Regional Fire Frequency. Peak components of total charcoal influx ($CHAR_i$) allow identification of fire events at local to regional scales (8, 10). First, $CHAR_i$ series were interpolated to constant time steps ($C_{\text{interpolated}}$), corresponding to the median temporal resolution of each record. Low-frequency variations in $CHAR_i$, namely $C_{\text{background}}$, represent changes in charcoal production, transport, sedimentation, mixing, and sampling. We therefore decomposed $CHAR_i$ into background ($C_{\text{background}}$) and peak (C_{peak}) components, using a locally defined threshold that identified charcoal peaks likely related to the occurrence of one or more local fires (i.e., “fire events” within *ca.* 1 km). The locally weighted regression was applied with a 500- to 800-y-wide window that maximized a signal-to-noise (peak-to-background) index and the goodness-of-fit between the empirical and modeled $C_{\text{background}}$ distributions (3). The residual series related to peaks were obtained by subtraction (i.e., $C_{\text{peak}} = C_{\text{interpolated}} - C_{\text{background}}$).

Consistent with theoretical evidence (10) and previous work (e.g., refs. 11–14), we assumed in a second step that C_{peak} was composed of two subpopulations, namely C_{noise} , representing variability in sediment mixing, sampling, and analytical and naturally occurring noise, and C_{fire} , representing significant peaks of charcoal inputs from local fires. For each peak, we used a Gaussian mixture model to identify the C_{noise} distribution. We considered the 99th percentile of the C_{noise} distribution as a possible threshold separating samples into “fire” and “nonfire” events. All statistical treatments were performed using the program CharAnalysis (P. E. Higuera; Moscow, ID).

From fire event dates extracted from C_{fire} over the past 7,000 y, we computed fire frequencies (FF), using a kernel smoothing regression function (Matlab open source function) that allowed a detailed inspection of time-dependent event frequencies (9). In practical terms, we used the Nadaraya–Watson kernel regression (15, 16) to calculate FF from fire event dates at a constant time step of 10 y from -50 to 7,000 cal yBP such as

$$FF = \frac{\sum_{i=-50}^{7,000} K_h(x_i - X_i) F_i}{\sum_{i=-50}^{7,000} K_h(x_i - X_i)}, \quad [S5]$$

where x_i are the dates where the interpolation of fire frequency took place (i.e., a series ranging from -50 to 7,000 with 10-y time steps), X_i are the dates of the fire events obtained by CHAR peak analysis, and F_i are the associated local frequencies computed such as $F_i = 1/(X_i - X_{i-1})$. K_h is the Gaussian kernel function such as

$$K_{(y,h)} = \frac{1}{\sqrt{2\pi h}} e^{-\frac{y^2}{2h^2}}, \quad [S6]$$

where h was the bandwidth ($h = 500$ y) selected by cross-validation, aimed at finding a compromise between large variance and small bias (small h) and small variance and large bias (large h). In the case of a pooled record of multiple lakes to reconstruct the regional fire frequency (*RegFF*), the values of FF were adjusted to the changing number of lake samples through time, using

$$RegFF_t = FF_t / n_t, \quad [S7]$$

where n_t was the number of sampled cores at time t . We assessed the significance of changes with the help of BCIs computed from confidence bands (90%) around *RegFF_t*. Detected trends in occurrence rate were confirmed by box-plot analysis carried out at each 1,000 y where the significance of differences between the FF_t from the 1,000-y periods was assessed using Wilcoxon’s rank sum test.

Fire-Season Length Reconstruction Using General Circulation Model Data Processing. We applied the method developed by Hély et al. (17) on climate simulations from the UK Universities Global Atmospheric Modeling Program (UGAMP) general circulation model (GCM) (18) to compute the fire-season length and to assess the fire risk throughout the Holocene.

We used outputs of new UGAMP climate simulation snapshots (1,000-y intervals) covering the last 7,000 yBP (19). Compared with the previous model simulations (17), the new simulation includes output data for the 2,000-yBP period, which were previously missing, it includes forcing from a prolonged presence of the residual Laurentide Ice Sheet, and it includes an improved way of handling the isostatic rebound that was previously less effective (19). As a consequence, the air temperature between 7,000 and 5,000 yBP is cooler in the new series (Fig. 2) compared with the previous series (17).

For each millennium, mean anomaly computation for air temperature (difference between UGAMP XK and 0K control periods) and precipitation (percentage of change between XK and 0K) were computed similarly to Hély et al. (17).

To increase the spatial resolution from the coarse GCM to a resolution more compatible for comparison with our local proxy records, temperature and precipitation UGAMP anomalies were applied to the means computed from the Climate Research Unit spatial grid TS 2.1 (Time Series at 0.5°) (20). As variance was not provided in the UGAMP simulations, we kept the present-day observed variance computed from the Climate Research Unit (CRU) datasets, namely from temporal variability (1901–2002) at 0.5° based on the Mitchell and Jones (20) dataset for precipitation and temperature.

Within each 0.5°-pixel grid, we used normal distribution for temperature and gamma distribution for precipitation (21) to obtain a 30-y time series for each month and with each variable, each specific monthly distribution being parameterized with the reconstructed monthly mean and prescribed variance (22). From this 30-y time series, we ran the Richardson (23) weather generator to reconstruct daily temperature and precipitation necessary for the computation of the Canadian Drought Code (24), which represents the moisture content of the deep organic matter (i.e., 18 cm thick and 25 kg·m⁻² dry weight, for a bulk density of 138.9 kg·m⁻³) that slowly fluctuates over the dry season. The reconstruction of the fire season length for each millennium was performed as in Hély et al. (17).

Seasonal and monthly means of climate outputs and solar irradiance were correlated to *RegBB*, *RegFF*, and the *FS* index, using Pearson's correlation coefficients. Significance was determined using permutation tests with correction for multiple comparisons (25, 26).

Effect of Mean Temperature on Mean Fire Size During Early-Season Burning. The influence of climate on mean annual size of early-season fires was examined as follows. Forest fire data from the Ministère des Ressources Naturelles et de la Faune du Québec were used for this study. The database contains information on the location, date of detection, size (hectares), and cause (lightning or human) of all fires recorded in the province of Quebec. The period covered by the data encompasses that during which systematic fire detection was done by detection planes. We considered lightning fires from 1973 to 2009 only. Spring fires were identified on the basis of their starting data, i.e.,

April to June. Mean annual fire size was then computed. Because the mean is a sensitive measure to skewness bias, we constrained our analyses to those years in which the number of fire samples was high enough for conducting bootstrap resampling. A satisfactory compromise was the use of a minimum threshold of $n = 10$ early-season fires per year, providing us with a sample size of 18 y for correlation analysis. Hence, those seasons in which the number of early-season fires was $n < 10$ fires were excluded from the correlation analysis.

Temperature and precipitation data for correlation analysis were those of the CRU TS 3.1 data, resolution of 1° latitude (1° longitude) (20). Monthly means of temperature and monthly totals of precipitation data for the domain encompassing 79.5°W–70.0°W and 48.5°N–52.5°N were averaged and correlated to the mean annual size of early-season fires. A high correlation was found using the month of June as a predictor variable for the mean annual size of early-season fires (Spearman's $R^2 = 0.46$, $P = 0.0015$, $n = 18$ y), which is consistent with the timing of the early-season area burned in these forests (ignition of 95% of the sampled fires took place after May 27th). No significant correlation was found for other months, and no significant correlation was found when using monthly precipitation totals as a predictor variable. No correlation was found between June temperature and number of fires. A sensitivity analysis conducted using a threshold of $n \geq 5$ fires also yielded a high correlation between mean fire size and June temperature (Spearman's $R^2 = 0.30$, $P = 0.0058$, $n = 24$ y); again, no correlation was found between June temperature and number of fires.

1. Stuiver M, Reimer PJ (1993) Extended ¹⁴C database and revised CALIB radiocarbon calibration. *Radiocarbon* 35:215–230.
2. Reimer PJ, et al. (2009) IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* 51(4):1111–1150.
3. Higuera PE, Brubaker LB, Anderson PM, Hu FS, Brown TA (2009) Vegetation mediated the impacts of postglacial climate change on fire regimes in the south-central Brooks Range, Alaska. *Ecol Monogr* 79(2):201–219.
4. Power M, et al. (2008) Changes in fire regimes since the Last Glacial Maximum: An assessment based on a global synthesis and analysis of charcoal data. *Clim Dyn* 30(7):887–907.
5. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* (Springer, New York), 4th Ed.
6. Bergeron Y, Gauthier S, Flannigan M, Kafka V (2004) Fire regimes at the transition between mixedwood and coniferous boreal forest in northwestern Quebec. *Ecology* 85(7):1916–1932.
7. Johnson EA, Gutsell SL (1994) Fire frequency models, methods and interpretations. *Adv Ecol Res* 25:239–287.
8. Higuera PE, Whitlock C, Gage JA (2011) Linking tree-ring and sediment-charcoal records to reconstruct fire occurrence and area burned in subalpine forests of Yellowstone National Park, USA. *Holocene* 21(2):327–341.
9. Mudelsee M, Börngen M, Tetzlaff G, Grünewald U (2004) Extreme floods in central Europe over the past 500 years: Role of cyclone pathway "Zugstrasse Vb" *J Geophys Res* 109:D23101.
10. Higuera PE, Peters ME, Brubaker LB, Gavin DG (2007) Understanding the origin and analysis of sediment-charcoal records with a simulation model. *Quat Sci Rev* 26(13–14):1790–1809.
11. Gavin DG, Hu FS, Lertzman K, Corbett P (2006) Weak climatic control of stand-scale fire history during the late holocene. *Ecology* 87(7):1722–1732.
12. Higuera PE, et al. (2008) Frequent fires in ancient shrub tundra: Implications of paleorecords for arctic environmental change. *PLoS ONE* 3(3):e0001744.
13. Ali AA, Carcaillet C, Bergeron Y (2009) Long-term fire frequency variability in the eastern Canadian boreal forest: The influence of climate vs local factors. *Glob Change Biol* 15(5):1230–1241.
14. Carcaillet C, et al. (2001) Change of fire frequency in the eastern Canadian boreal forests during the Holocene: Does vegetation composition or climate trigger the fire regime? *J Ecol* 89(6):930–946.
15. Watson GS (1964) Smooth regression analysis. *Sankhyā Ser A* 26(4):359–372.
16. Nadaraya EA (1965) On non-parametric estimates of density functions and regression curves. *Theory Probab Appl* 10(1):186–190.
17. Hély C, et al. (2010) Eastern boreal North American wildfire risk of the past 7000 years: A model-data comparison. *Geophys Res Lett* 37:L14709.
18. Hall N, Valdes P (1997) A GCM simulation of the climate 6000 years ago. *J Clim* 10(1):3–17.
19. Singarayer JS, Valdes PJ (2010) High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quat Sci Rev* 29(1–2):43–55.
20. Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climatol* 25:693–712.
21. New MG, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Clim Res* 21:1–25.
22. Ramstein G, et al. (2007) How cold was Europe at the Last Glacial Maximum? A synthesis of the progress achieved since the first PMIP model-data comparison. *Climate of the Past* 3(2):331–339.
23. Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour Res* 17(1):182–190.
24. van Wagner CE (1987) *Development and Structure of the Canadian Forest Fire Weather Index System* (Canadian Forestry Service, Ottawa).
25. Groppe DM, Urbach TP, Kutas M (2011) Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology* 48(12):1711–1725.
26. Robinson A (2007) Randomization, bootstrap and Monte Carlo methods in biology. *J R Stat Soc Ser A Stat Soc* 170(3):856–856.

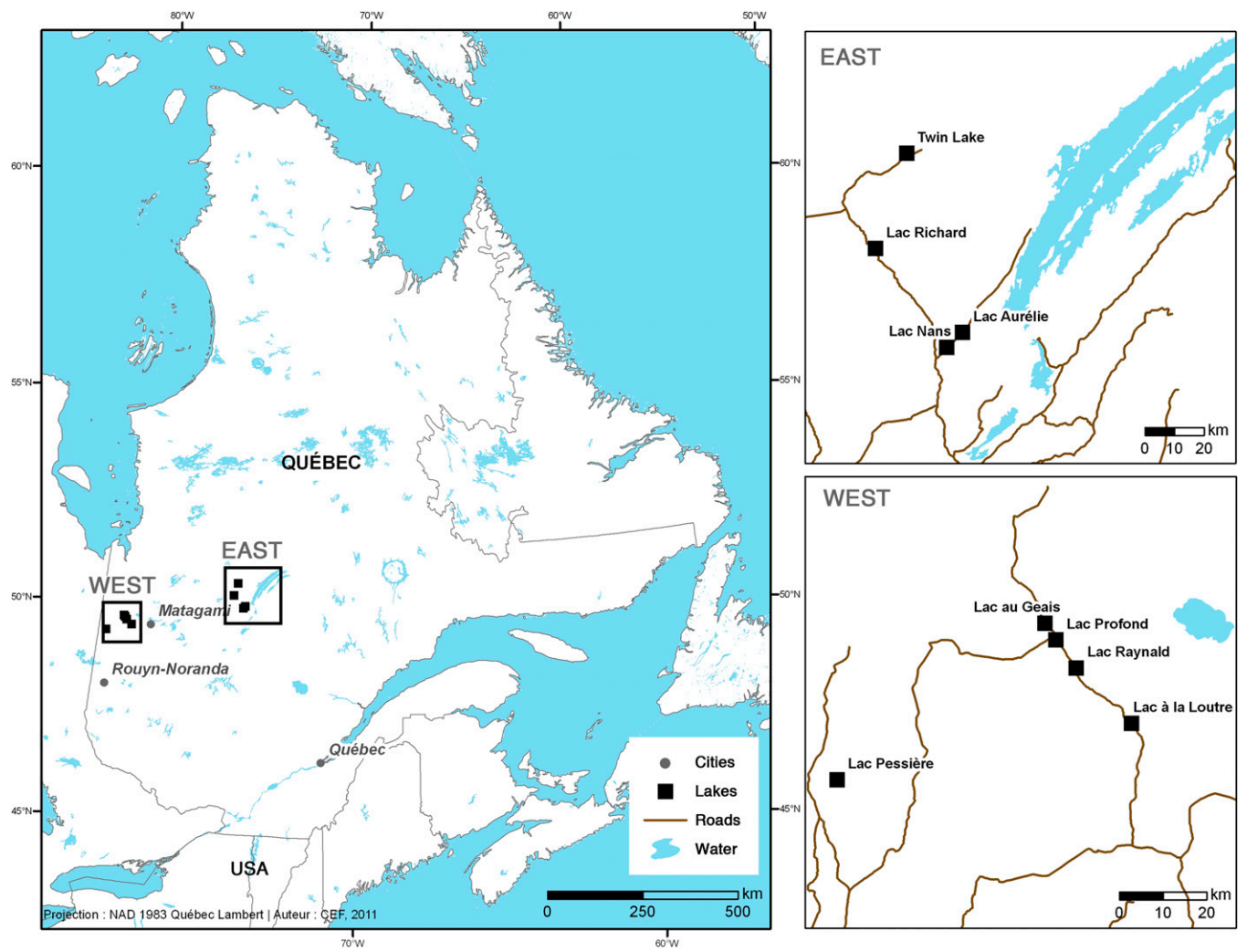


Fig. S1. Site location map.

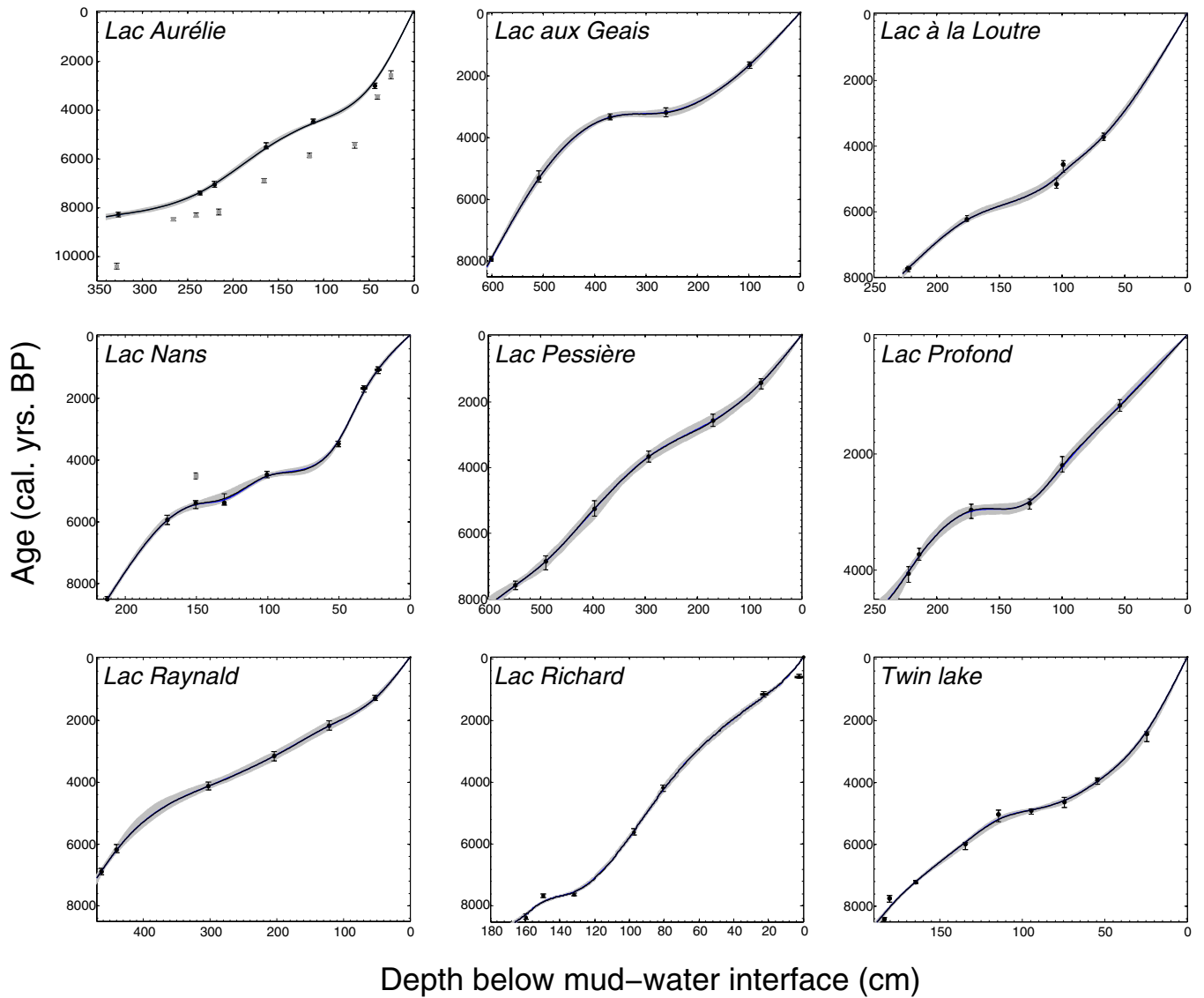


Fig. S2. Age-depth models for the cores from the nine studied lakes. Black dots represent the calibrated radiocarbon dates with associated error (vertical bars), the black line represents the spline-based age-depth model with resampled 95% confidence intervals (gray area). Dates not considered in the models appear as gray dots (Nans and Aurélie lakes). For the Lac Aurélie, dates were rejected on the basis of the presence of biogenic carbonate (sediment rich in *Chara oogonium*).

