Tree biomass reconstruction shows no lag in post-glacial afforestation of eastern Canada.

Blarquez, Olivier and Aleman, Julie C.

Abstract: Forest ecosystems in eastern Canada are particularly sensitive to climate change and may shift from carbon sinks to carbon sources in the coming decades. Understanding how forest biomass responded to past climate change is thus of crucial interest. But past biomass reconstruction still represents a challenge. Here we used transfer functions based on modern pollen assemblages and remote sensed biomass estimation to reconstruct and quantify, for the last 14 000 years, tree biomass dynamics for the six main tree genera of the boreal and mixed-wood forests (Abies, Acer, Betula, Picea, Pinus, Populus). We compared the mean genera and total biomass to climatic (summer temperatures and annual precipitation), physical (CO₂, insolation, ice area) and disturbance (burned biomass) variables, to identify the potential drivers influencing the long-term trends in tree biomass. Tree biomass was for most genera related with summer temperature, insolation and CO₂ levels; Picea was the exception and its biomass also correlated with annual precipitation. At the onset of the Holocene and during the Holocene Thermal Maximum (c. 10000-6000 BP), tree biomass tracked the melting of the Laurentide Ice sheet with high values (>50 t.ha⁻¹ and a total of 12 Pg). These values, in the range of modern forest ecosystems biomass, indicate that trees were probably able to survive in a periglacial environment and to colonize the region without any discernible lag by tracking the ice retreat. High biomass at the beginning of the Holocene was likely favoured by higher than present insolation, CO₂ levels higher than during the Last Glacial Maximum, and temperature and precipitation close to present day levels. Past tree biomass reconstruction thus brings novel insights about the drivers of postglacial tree biomass and the overall biogeography of the region since the deglaciation.

Keywords

Tree biomass, pollen, Last Glacial Maximum, Holocene, Modern Analogue Technique, Climate.

1. Introduction

In eastern Canada, forests are particularly sensitive to climate change, and numerous factors may contribute to reversing their carbon sink capacity into a source, as for example increased disturbance frequency by fire (Mack et al., 2011) or biotic agents (Hicke et al., 2012), increased drought frequency (Ma et al., 2012), or increased permafrost thawing and the associated emissions from soil micro-organisms at the forest-tundra ecotone (Schuur et al., 2009). For example, carbon emissions...
from fires in northern ecosystems, notably in tundra and boreal forest, represent a significant proportion of global emissions (Bradshaw et al., 2009). Reconstructing the temporal dynamics of forest biomass is crucial to understand the processes that influence carbon storage and to help forecast forest responses to ongoing global changes. Additionally, forest ecosystems are composed of long-lived organisms, and the mechanisms underlying their dynamics take place at a scale spanning centuries to millennia (e.g., forest succession, range expansions, Feurdean et al., 2013, Bergeron et al., 2004). To put into perspective the temporal dynamics of forest biomass regarding climate and natural forcing, reconstructions gain to be performed at the millennial time scale. To date, the main source of information that could be used to infer past biomass comes from the analysis of bio-proxies, such as pollen (Seppä et al., 2009) or plant macro-remains (Blarquez et al., 2012), extracted from sedimentary archives, or the analysis of historical vegetation surveys (Rhemtulla et al., 2009).

Transfer functions could be used to relate paleoecological proxies accumulation in sediments with environmental conditions (e.g. climate, pH, etc.) or communities features (e.g. vegetation cover, biome type). Those functions are based on the principle that biological samples (e.g. pollen) with similar taxa composition were produced by plant communities with similar compositional and structural characteristics. Among those functions the modern analog technique (MAT) is one of the most widely used in paleoecological studies (Overpeck and Webb, 1985). MAT consist first to (i) calibrate modern pollen assemblages against modern vegetation features, then (ii) modern pollen samples are compared to fossil samples, (iii) the closest analogues based on compositional similarity are chosen and finally (iv) the fossil pollen samples are assigned the vegetation feature associated with its closest modern analogs. This technique has been successfully used to reconstruct tree cover in North America since the Last Glacial Maximum (LGM 21,000 BP) with MAT calibrated using pollen assemblages and satellite tree cover measures (Williams, 2002, Williams et al., 2011).

The calibration step generally involves the comparison of modern samples with standing vegetation by using land surveys (Seppä et al., 2009), trapping protocols (Blarquez et al., 2012) or remote sensing data (Williams, 2002). Here the availability of quantified and homogeneous vegetation measures at the same spatial scale than modern pollen databases is crucial (Whitmore et al., 2005). Such data have been recently available for living tree aboveground biomass covering Canada in the form of maps of specific tree biomass estimations at a 250 x 250-m resolution for a vegetation gradient that ranges from the temperate forest to the tundra (Beaudoin et al., 2014). Before relating pollen assemblages to modern biomass certain precautions must be taken and assumptions be made. Notably, the accumulation of pollen in sediments is affected by numerous taphonomical biases, including differential productivity and dispersal of taxa, as well as the spatial scale for which pollen are representative of the population actually producing them (Bradshaw and Webb III, 1985). These uncertainties are also associated with sedimentary archive type and characteristics such as the deposit environment, the basin size, the sediment accumulation rate or the choice of an age-depth model. Here we used pollen relative abundances (i.e. percentage data) because they help standardize the data between variable sedimentary sites (Williams, 2002). Pollen accumulation rates should allow for absolute estimates of pollen concentrations and so are superior in theory to reconstruct biomass (Seppä et al., 2009), but in practice they are too sensitive to taphonomic biases and no modern pollen database expressed in accumulation rate is available (Whitmore et al., 2005).

Our objectives were thus to use MAT to relate modern tree biomass to modern pollen assemblages from Whitmore et al. (2005); and to compare modern and fossil pollen assemblages expressed in relative abundances for reconstructing the past biomass of ecosystems. We focused on eastern Canada (i.e., east of 100°W) because (i) the dominant tree forest species that span the area are replaced with non-analogous western and mountain species west 100°W (e.g., Abies balsamea, replaced with Abies grandis or Abies lasiocarpa), which would thus require a regional split of pollen types to ensure MAT accuracy (Williams and Shuman, 2008), and (ii) because of the Canadian distribution of the calibration dataset (Beaudoin et al., 2014). These quantified data we obtained are crucial to understand the bio-
geography of the region and could provide useful information for models linking vegetation dynamics to climate and/or disturbance. Finally, we compared tree biomass dynamics to climatic and physical variables in order to highlight the potential drivers of the afforestation process and evaluate its temporal variability.

2. Material and Methods

2.1. Modern Analogue Technique

The modern pollen samples that have been used for constructing the transfer functions were selected from the North American Surface Sample Dataset (Figure 1, Whitmore et al., 2005). Before the analysis, all samples were standardized to consider, at minimum, the genus level and the family level when necessary. For example, when for certain taxa the determination level between analysts was not homogeneous enough for classification to the genus level we aggregated those taxa at the family level (Williams and Shuman, 2008). Fifty-three genera and 28 families were considered (Table S1).

We used the Modern Analogue Technique (MAT) based on Squared-Chord Distance (SCD) to establish transfer functions between modern pollen assemblages and tree biomass (Overpeck and Webb, 1985). We used the tree biomass of the six forest tree genera that contributed the most to the total biomass above c. 40°N, i.e., Picea, Pinus, and Abies for needle-leaved trees and Acer, Betula, Populus for broad-leaved trees. Those genera are representative of the Canadian boreal, mixed boreal and Laurentian forests, and we limited our interpretation to those ecosystems. We used the biomass maps at 250 x 250-m resolution produced by Beaudoin et al. (2014) and based on a k Nearest Neighbour (kNN) approach and MODIS imagery. According to Beaudoin et al. (2014) biomass estimates error is greater in mountain regions, in low biomass areas and in regions were calibration data (i.e. aerial photographs) for the kNN procedure are sparse. Those situations do not generally apply for modern pollen assemblage sites located in the boreal and temperate forest (Fig. 1) and the spatial aggregation of biomass pixel values to 1 km² and beyond for all sites (see below) likely improved biomass estimates accuracy (Beaudoin et al., 2014). To obtain genera biomass maps we summed the biomass of several species: Picea represented the sum of P. glauca and P. mariana; Pinus that of P. banksiana, P. strobus and P. resinosa, and Abies represented Abies balsamea solely. For broadleaved trees, Acer was obtained by summing the biomass of A. rubrum, A. saccharum, A. saccharinum, A. pensylvanicum, and A. spicatum; Betula that of B. papyfera and B. alleghaniensis and finally Populus that of P. tremuloides and P. balsamifera. We selected the modern pollen samples located in Canada, east of 100°W (Figure 1).

Before calculating modern tree biomass estimated from remote sensing around modern pollen samples, we evaluated the optimal distance for extracting biomass values by replicating MAT calibration step with increasing distances such as radius=1,5,15,...,150 km (Figure 2). For each set of MAT replicates we evaluated the leave one out cross validation determination coefficient and its differential. The choice of an optimal distance was based on a compromise between high r², a small distance from the modern pollen sample and a r² differential approaching or equal to 0, which indicates that increasing the radius did not significantly increase MAT predictive power (Figure 2).

We used these distances for compiling the average biomass for each tree genus around each modern pollen sample. We evaluated the overall performance of the transfer functions by looking at predicted vs observed biomass distribution, associated determination coefficients and prediction uncertainty expressed by the Root-Mean-Square Error of Prediction (RMSEP, Figure 3). The number of analogues needed to reconstruct past biomass (k) for each genus was evaluated by bootstrap re-sampling of the modern dataset, split into training and test sets (Simpson, 2007). Although the choice of k is generally not strongly biased despite being determined post hoc from the training data, we used an independent optimization set alongside the usual training and test sets to avoid any bias in its choice (Telford...
et al., 2004). Before applying MAT, we used the Receiver Operating Characteristic (ROC) curve approach (Gavin et al., 2003). Each sample in the modern dataset was assigned to a single vegetation zone according to Ramankutty and Foley (1999). The SCD between each samples and its \((k)\) closest analogues within each vegetation zone was compared with the \((k)\) closest SCDs from other vegetation zones. This approach enabled to calculate the critical SCD value above which two samples cannot be selected as true analogues (Gavin et al., 2003). The MAT transfer functions were then applied to fossil pollen assemblages extracted from the Neotoma database (http://www.neotomadb.org) that were standardized to the same taxonomic levels we used for the modern dataset (see Table S1 for taxonomy and Table S2 for the full citation list of the Neotoma sites, Grimm et al., 2013). We selected the pollen records (or sites) from the Neotoma database that include at least one geo-chronological control point (i.e., \(^{14}C\) or \(^{210}Pb\) dating, tephra layer, etc.) for every 2500-year interval.

One hundred fifty two (152) sites were selected; the age-depth models of 133 were corrected according to the most up to date chronologies calculated by Blois et al. (2011). Eighteen sites, whose chronologies were expressed in radiocarbon years before present (\(^{14}C\) BP), were converted to calibrated years (cal BP) using the IntCal13 calibration curve (Reimer et al., 2013), and the procedure developed by Grimm (2008). The chronology of one site, expressed in calibrated years BP, was left unmodified. Table S2 lists the sites, the transformation applied on the age-depth models, the median time difference between original and corrected age-depth models, the location and the original references for the datasets.

2.2. Biomass mapping

We used weighted spatio-temporal interpolation to produce gridded maps of reconstructed biomass values from selected pollen cores. A spatial grid of 50 km square pixels was used to interpolate biomass values for 1000 years intervals since 14000 BP such as \(\text{Ages} = 14000, 13000, \ldots, 0\text{BP}\). Spatial interpolation involved searching sites with biomass values located at a horizontal distance of 300 km and a vertical distance of 500 m from each grid cell centre. The temporal interpolation involved identifying biomass values within a temporal window of 500 years before and after each key date. Biomass values were then weighted using a tricube weighting function by considering their spatial distance to the grid centre and temporal distance to the key date. This approach, which weights samples according to their spatio-temporal location, also down-weights pollen sites (and thus reconstructed biomass values) that are poorly sampled (Williams et al., 2004). For each pixel, we used the weighted pollen biomass estimates to calculate the mean biomass per hectare (in \(t.ha^{-1}\)) for the six considered genera. We summed the values for those six genera to obtain the mean total biomass per pixel.

2.3. Climatic and physical variables

Laurentide Ice Sheet maps from Dyke (2004) were first calibrated to calendar ages (cal BP) using the IntCal13 (Reimer et al., 2013) calibration curves. Those maps were then used to constrain the biomass estimate maps and to compile the total area of the Laurentide Ice Sheet that was compared to the biomass values. We used paleo-climatic simulations from the HadCM3 GCM (Singarayer and Valdes, 2010) and from the CCSM3 GCM (Liu et al., 2009) obtained from http://ccr.aos.wisc.edu/resources/data_scripts/ to compare biomass dynamics to mean summer temperatures (June, July and August) and mean annual precipitation anomalies. Those simulations consists of climatic averages at 1000-year intervals (i.e., maximum temporal resolution available) covering the last 14000 years at a spatial resolution of 2.5° in latitude by 3.75° in longitude. For each experiment and each millennium interval, anomalies for air temperature (the difference between a given millenium and the pre-industrial (AD c. 1750) period) and precipitation (the percentage of change between a given millennium and the pre-industrial period) were computed and downscaled using the CRU 3.10 and CRU 3.0 datasets for the HadCM3 and CCSM3 respectively. For each millennium, we compiled the mean anomalies of summer temperatures (June, July, August in °C) and mean annual precipitation (in
mm.day-1) for pixels matching the biomass estimates maps and bootstrapped confidence intervals. We used the δ18O of the NGRIP ice core record as an additional proxy for regional temperatures (Andersen et al., 2004), ice core CO2 values from the Epica Dome C coring as a proxy of past CO2 atmospheric content (Monnin et al., 2004) and insolation values at 45°N calculated using orbital parameters (Berger and Loutre, 1991). For each millennium since 14000 BP, we calculated the mean values of those proxies in 1000-year time windows (centred on millennia). Biomass burning trends were evaluated based on sedimentary charcoal composite curves. The approach is described in Blarquez et al. (2014b) and followed the data transformation explained in Power et al. (2008) involving a three step transformation of charcoal data (min-max, Box-Cox and Z-Score successive transformations). The compositing approach followed Daniau et al. (2012) and Marlon et al. (2013). The sites were obtained from the Global Charcoal Database version 3 (http://paleofire.org).

2.4. Statistical analyses

We used the biomass maps for each genus to calculate millennia mean and total biomass values. We summed millennia genus biomass values to obtain the total mean biomass value at each millennium. We used the Least Absolute Shrinkage and Selection Operator (LASSO) regression method (Tibshirani, 1996) to select climatic and physical variables that most strongly influence the total and each genera biomass dynamics. We used a 10 fold cross validation approach to find the optimal λ value of the LASSO regression that was then used to select the predictors of the model different from zero. Because the LASSO procedure is not invariant to linear transformations of the predictors, the physical and climatic variables were rescaled using Z-Scores prior to the analysis. This rescaling step further enabled to rank and to compare the predictors based on the value of their regression coefficients, a larger coefficient meaning that the predictor variable was more explanatory. The temporal trend in mean biomass was evaluated by fitting a LOWESS curve with a 200 year window half width to the reconstructed biomass for all sites (Blarquez et al., 2014b, 2015).

3. Results

3.1. MAT accuracy

MAT was based on a set of 594 modern pollen samples distributed over the Canada, east of 100°W. The distance that satisfied best the compromise between high r2, a small distance from the modern pollen sample and r2 differential approaching or equal to 0, were 17 km for Abies, 31 km for Acer, 25 km for Betula, 27 km for Picea, 20 km for Pinus, and 26 km for Populus. For Populus although d(r2) was 0 at 10 km the 26 km distance significantly increased r2. These distances are consistent with empirical studies of pollen source-area (Bradshaw and Webb III, 1985) but are much higher compared with Sugita (1994), notably for Abies. However, we must consider that those distance are not representative of the direct relationship between a particular taxa and its landscape biomass but rather involve pollen assemblage relationship with genus landscape biomass expressed through MAT, which largely explain those longer distances. These distances are however in agreement with studies that aimed at relating remote sensed woody cover with pollen assemblages (Williams, 2002). The examination of predicted vs observed reconstructed biomass for each genus provides a coherent pattern indicating that the biomass values were well modelled (Figure 3). For most genera the error between modelled and observed biomass was inferior to 5 t.ha−1, but the error was inherently higher for Picea due to its high biomass values notably in the boreal coniferous forests (Figure 1). Additionally, maps comparing biomass predicted using MAT and fossil pollen samples from the 1900-2010 period and current biomass showed that the error could be important in areas that have experienced large anthropogenic influence during the last century (for example in the St Laurence valley for Acer or Betula, Figure S1). The MAT determination coefficients were always above 0.5, and the Root-Mean-Square Error of Prediction (RMSEP) was below 3 t.ha−1 for most genera. This indicates that the reconstructions are
accurate enough for reconstructing general tendencies of biomass dynamics at large spatial and temporal (millennia) scales. The number of analogues chosen by the optimization procedure ranged from 3 for *Populus* to 12 for *Picea* (Figure 3). The best prediction was for *Betula* that showed both a high determination coefficient (0.82) and a low RMSEP (2.03 $t. ha^{-1}$). For *Picea*, the RMSEP was high, with a value of 6.88 $t. ha^{-1}$, suggesting less certain reconstruction. But this value must be compared to the overall biomass range of the genus (0-60 $t. ha^{-1}$), which was almost twice that of all of the other genera. For all of the genera, the MAT tends to underestimate biomass values when observed biomass was above c. 10 $t. ha^{-1}$ (Figure 3). The SCD critical value obtained from the ROC curve approach above which samples cannot be selected as true analogues was 0.14 SCD units. This value resulted in the exclusion of samples without analogues in the modern dataset mainly during the period before 10000 BP where the number of samples without analogues was as high as 50% (Figure 4). After 10000 BP and during the Holocene this number stayed below 25% (Figure 4).

### 3.2. Spatial biomass dynamics

Before 14000 BP, the number of pollen sites was not sufficient enough for tree biomass reconstruction, due to the presence of the Laurentide Ice sheet (Figure 4). After 14000 BP, isolated sites began to accumulate sediments and thus pollen, enabling the reconstruction of biomass using MAT and the production of interpolated maps with more confidence since c. 10000 BP (Figure 4). Generally, pollen core locations are well distributed across the territory and using the spatial weighting procedure it was possible to produce maps covering the entire territory (excluding the north-western area corresponding to North Ontario and East Manitoba, where no pollen cores were available).

After 13000 BP, *Picea* biomass increased up to its maximal level (with values above 19 $t. ha^{-1}$) in the direct vicinity of the ice sheet (Figure 5). Before 11000 BP, hotspots of *Picea* biomass were located east of the Great Lakes, but the overall area supported high *Picea* biomass with spatially continuous values above 11 $t. ha^{-1}$. At 9000 BP, a biomass hotspot was located between the Agassiz and Objiway Lakes. After pro-glacial lakes disappeared, this hotspot tracked the melting of the remaining ice sheet eastward. From 5000 BP to 10000 BP, the biomass values for *Picea* were high across the region, with two notable centres of very high values located (i) east of Winnipeg Lake in Manitoba and Western Ontario and (ii) in the Quebec Côte Nord region and southern Labrador.

Before the Holocene period (starting c. 11700 BP), *Abies* biomass values were low, below 5 $t. ha^{-1}$ (Figure 5). Then, at the period centred around 11000 BP, (11500-10500 BP) *Abies* biomass started to increase, following a westward route from Nova Scotia, where high biomass values were recorded (between 8 and 29 $t. ha^{-1}$). During the Holocene, the area that displayed the highest values of *Abies* biomass remained spatially constant and localized in the northern part of the St. Lawrence region, Quebec Côte-Nord region and the Canadian Maritime provinces (New Brunswick, Nova Scotia).

High *Pinus* biomass values started to be recorded at 10000 BP in an area that is bounded in the north by the pro-glacial lakes, Agassiz and Objiway, and South by the Great Lakes. Up to the present day, the spatial distribution of the *Pinus* biomass was characterized by high values (> 7 $t. ha^{-1}$) located in the south-western Ontario and Manitoba (Figure 5). The most notable *Pinus* biomass feature were the westward expansion between 12000 and 10000 BP and a shrinking of its spatial distribution since 8000 BP that accelerated from 3000 BP (Figure 5).

The spatial distribution of *Betula* biomass was largely similar with that of *Abies*. However the localization of high *Betula* biomass values (> 11 $t. ha^{-1}$) was generally limited to the east of the St. Lawrence River and never span Newfoundland (Figure 5).

The spatial pattern of *Populus* biomass was marked by a westward expansion since 13000 BP. This expansion peaked at 8000 BP (Figure 5). During the 8000 BP period, particularly high *Populus* biomass values (from 9 to 35 $t. ha^{-1}$) were found north of the Great Lakes in the area left by the draining of Agassiz and Objiway Lakes. *Populus* also expanded in the East (East of the St. Laurence river, Nova Scotia and New Brunswick) after 9000 BP but high biomass values disappeared from that area after
6000 BP. After that period and up to the present day, the area of high *Populus* biomass shrank and remained located west of the St. Lawrence River, with a smaller area displaying high values located in Ontario and Manitoba between the Nigipon and Winnipeg Lakes (Figure 5).

*Acer* biomass values became high (i.e., between 12 to 42 t.ha$^{-1}$) only after 8000 BP, mainly in the east of the St. Lawrence region, in areas that correspond to the current New Brunswick, Nova Scotia and Quebec’s Gaspé Peninsula. *Acer* spatial distribution then remained relatively constant and confounded with its present day distribution. For *Acer*, but also for *Betula* we observed, during the last 500 years, a shrinking in their eastern distribution with a decrease of their biomass east of the St. Lawrence River and in the Maritimes provinces.

The overall total biomass, i.e. the pixel sum of the six studied genera, started to increase, with very high values at approximately 10000-9000 BP, in an area that corresponds to the current Maritime provinces and Quebec (Figure 5). Then, areas of high biomass ($>50$ t.ha$^{-1}$) spanned the north of the Great Lakes at 8000 BP mainly because of *Populus* and *Pinus* expansion. After 8000 BP, a contiguous belt of high biomass values ($> c. 43$ t.ha$^{-1}$) matched the spatial distribution of the current boreal and boreal mixedwood forests. This area was split at its centre in an area that corresponds to the south centre of Québec, favoured by the highly defined east-west pattern for the studied genera (east distribution for *Acer*, *Betula* and *Abies* and west distribution for *Pinus* and *Populus*, Figure 5).

### 3.3. Ranking the influence of physical and climatic processes

Summer insolation, GCM simulated summer temperatures and atmospheric $CO_2$ content were the variables that had the strongest correlation to the total mean tree biomass (Figure 6 and Table 1). The total mean biomass was maximal during the period corresponding to the Holocene Thermal Maximum (HTM, c. 10000-6000 BP Kaufman et al., 2004). The LASSO procedure indicated that the genera dynamics followed this general pattern, i.e. a millennial dependence upon climatic condition (temperatures), solar energy expressed by insolation, and a positive relationship with the atmospheric $CO_2$ content. All genera appeared to be positively influenced by summer insolation and temperatures (expressed by NGRIP $\delta^{18}O$ or simulated by GCMs), especially *Pinus*. *Picea* biomass appeared to depend also on annual precipitation contrary to the other genera for which precipitation did not appear to be limiting or which expressed a negative relationship to precipitation. *Picea* relationship to temperatures is unclear, has a negative coefficient with $\delta^{18}O$ and a positive one with summer simulated temperature (Figure 6, Table 1). Biomass burning appears related only with *Picea* and *Abies*, but this variable always has low coefficients compared to the other variables (Table 1).

### 4. Discussion

#### 4.1. The spatio-temporal pattern of postglacial biomass construction and its drivers

The reconstructed spatio-temporal pattern of *Picea* biomass dynamics was coherent regarding to the known history of the genus and its current status, particularly *Picea mariana* that represents the highest contributor to the genus’ biomass in the boreal forests. Moreover, this genus is the most abundant in the studied region, which corresponds to the boreal and mixed-wood forest, at least since the LGM. Interestingly, the event known as the *Picea* decline that is supposed to span between 11000 BP and 9000 BP is not apparent from our reconstructions (Figure 5 and 6). Blois et al. (2011) precisely dated this *Picea* decline, by Bayesian change-point analysis, around 11500 BP and 10100 BP, from southern to northern Canadian sites respectively. Although we cannot rule out *Picea* decline from occurring in eastern Canada, we can suspect that its decrease in pollen percentage may be the result of the proportional increase in other trees, notably broad-leaved trees that expanded their biomass during the early Holocene (Figure 6). Watts (1979) showed that the *Picea* pollen decline was not accompanied with a decline in *Picea* needles macrofossils and was thus likely a statistical artefact due to higher *Pinus* pollen production. It is then unsurprising that uncalibrated bio-proxies dynamics may be misleading.
and germinates on bare mineral soils. It has been shown that it was difficult to absolutely quantify its presence in the past landscapes. Those tundra-like plant communities were characterised by a higher diversity (Blarquez et al., 2014b), contained assemblages without ana-

contrary to Betula, Populus pollen is known to be poorly preserved in fossil records, making it difficult to absolutely quantify its presence in the past landscapes. Populus genus gathers pioneer trees (P. tremuloides and P. balsamifera) that produces large quantities of seeds, forms large clone colonies and germinates on bare mineral soils. It has been shown that Populus tremuloides growth is favored by higher June temperatures and a low thickness of the soil organic layer (Gewehr et al., 2014). Between 9000 BP and 8000 BP, the ice front continued to recede northwards, Lake Ojibway and Lake Agassiz drained (resulting in the 8200 cal BP climatic event) and a large lakebed containing glaciolacustrine deposits of sands, silts and clays remained, thus creating favourable conditions for the invasion and growth of Populus, i.e. bare grounds and climatic conditions characteristic of the HTM (Figure 6). Then, during the 8000 BP period, Populus attained its maximal spatial coverage and biomass with values up to c. 35 t.ha$^{-1}$ (Figure 5 and 6). This result is in agreement with Comtois and Payette (1984) who concluded that the forest development phase in Southern Québec was accompanied by Populus, but these authors suggested the presence of open landscapes, in opposition with our biomass reconstruction that shows values above c. 50 t.ha$^{-1}$ as early as 10000 BP (Figure 5). These values are characteristic of forest and close to the current estimates for the boreal forest obtained from allometric equations (41.8±2 t.ha$^{-1}$, Botkin and Simpson, 1990) or remote sensed data (73 t.ha$^{-1}$, Ranson et al., 1997). Growing evidence suggests that trees were able to survive and grow in periglacial conditions, i.e. directly at the proximity of glaciers, during the last glaciaion in Central Canada (Bélanger et al., 2014), in Alaska (Brubaker et al., 2005) or in Scandinavia (Kullman, 2002). Moreover, the existence of an initial treeless tundra between the last deglaciation and the invasion of trees is still controversial (Liu, 1990). Macrofossil evidence has suggested the local presence of trees at the ice margin as early as 12000 BP (Jackson et al., 1997) and some pollen studies indicated that the proximity of ice did not necessarily support tundra-like vegetation, but rather forests, at the onset of the deglaciation (Richard et al., 1982). We do not rule out the presence of transient tundra-like vegetation before 10000 BP that was interpreted from numerous pollen records (Ritchie, 2004). Those tundra-like plant communities were characterised by a higher diversity (Blarquez et al., 2014b), contained assemblages without ana-

The mean biomass of the two other conifer genera, Abies and Pinus, followed a linear increase from 14000 to 11000 BP. During the HTM, both their mean biomass and spatial extent were maximal, potentially favoured by climatic conditions (temperatures and insolation, Table 1). This agrees with previous evidence suggesting that Pinus, notably Pinus banksiana/resinosa, was more abundant during the early Holocene period (before 7000 BP, Fuller, 1997, Liu, 1990). The spatial pattern of Pinus biomass clearly showed a westward expansion of the high biomass areas that were previously located in areas filled with the pro-glacial lakes. The temporal pattern of this expansion is thus coherent with the Pinus decline observed across Southern Ontario (Fuller, 1997, Liu, 1990) and in the clay belt (Carcailllet et al., 2001).
logues in the modern dataset (Williams and Jackson, 2007) and so were seemingly excluded by the ROC screening in our analysis (Figure 4). This would eventually account for overinflating the mean biomass before 10000 BP, but including no-analogue samples in the reconstruction did not result in a mean biomass lowering (results not shown, Figure 6). Then, the existence of tree populations at the ice margin likely explain why biomass construction tracked the ice retreat and saturated so quickly. This hypothesis is among the only to also explain Reid’s paradox, i.e. why post-glacial trees migration rates are generally found to be faster compared with tree dispersal capacities (Feurdean et al., 2013).

The only genus whose biomass increase and postglacial migration appeared to be delayed is *Acer*. The increase in *Acer* biomass really began only during the Holocene and particularly after 6000 BP when populations of *Acer saccharum* expanded in the eastern regions (Muller and Richard, 2001) (Figure 5). After this colonization phase, *Acer* became the second most important taxa in terms of biomass, reaching values up to 42 t.ha$^{-1}$ (Figure 6). This biomass pattern is in agreement with Muller and Richard (2001) results, which indicated that the dominance of *Acer* in southern Québec resulted from a dominance shift and, thus, from competition processes rather than from a migratory invasion.

Total tree biomass attained values ranging between 50-70 t.ha$^{-1}$ as early as the beginning of the Holocene period, in areas located in the direct vicinity of the ice margins (Figure 5). Maximum genera biomass was attained during the HTM probably because of favorable climatic conditions but likely mainly because of higher than present insolation that led to higher energy released to the ground, then available for photosynthesis and tree growth. During the HTM, CO$_2$ levels were comparable to the Holocene long term trend (> 250 ppmv) and thus likely not limit biomass construction for most genera (Figure 6 Gerhart and Ward, 2010). The interplay between CO$_2$ and vegetation is not trivial however, particularly during the LGM where low CO$_2$ may have had strong physiological effects (Gerhart and Ward, 2010) and at the beginning of the Holocene, because fires in the boreal region (whose fuel are dependent upon vegetation biomass, Blarquez and Carcailliet, 2010) may have induced the postglacial increase in CO$_2$ levels (Carcailliet et al., 2002).

Our results may have strong implications for understanding how tree biomass may respond to ongoing climate change. For example, the most pessimistic ‘representative concentration pathway’ developed by the IPCC, RCP 8.5, is predicting an increase of CO$_2$ up to 1300 ppm and a temperature increase of 4°C by 2100 (Van Vuuren et al., 2011). CO$_2$ fertilization may contribute to increase tree growth rate (Mousseau and Saugier, 1992) and water efficiency (Keenan et al., 2013), and therefore would stimulate forest productivity and a higher tree biomass compared to present day. Higher temperature can also increase forest productivity, such as what happened during the HTM, but its role can be offset by the CO$_2$ increase that would hence stabilize growth rates (Tjoelker et al., 1998). Other factors may compensate for potential increase in tree biomass such as natural and anthropogenic disturbances. Fire frequency is projected to increase in the future decades (Flannigan et al., 2005), as pest outbreaks will (Dale et al., 2001), such that those disturbances by consuming biomass may offset the biomass increase stimulated by favourable climatic conditions. Potentially, species would tend to extend their northern range in order to track the warming, but it is not yet certain that boreal and temperate tree species will attain the migration rates required to cope with climatic changes (Looke et al., 2009). Indeed few evidences suggest northward migration of trees in eastern Canada (Caccianiga and Payette, 2006). Moreover, the environmental conditions that these species would meet by extending their range are unknown, notably the fire-vegetation feedback (Girardin et al., 2013), as well as about potential resilience of ecosystems under no-analogue climates (Williams and Jackson, 2007, Gauthier et al., 2015). Likewise, wood production and management in the commercial forests must be taken into account when predicting future biomass and ecosystem resilience to global changes (Gauthier et al., 2015).

The LASSO procedure results should generally be taken with caution to inform about the general temporal trends in biomass and its relationship to explaining factors, because we must assume that the role of temperature, precipitation, insolation or CO$_2$ have varied through time and space. Moreover,
we found that insolation and $CO_2$ were the best predictors of biomass trends, while the role of temperature is less clear. This can be a surprising result that may be explained by the differences among proxies uncertainty. Indeed, both insolation and $CO_2$ are estimated precisely from orbital parameters or ice cores but the temperature and precipitation data are simulated using general circulation models, which can be the source of inaccuracies. Temperature and precipitation may thus have played a more important role in biomass construction than highlighted by the regression model alone (Table 1). Interestingly biomass burning trend such as expressed by the charcoal composite curve is always a poor predictor of tree biomass. The few weak relationships that have been found concerned conifers, for example a positive one with *Picea* that is generally known to fuel fires (Johnson, 1992) and a negative one with *Abies* and thus *Abies balsamea* that is known for being the boreal conifer the least adapted to fire (Table 1). It has been shown that biomass burning is mostly a result of vegetation features through fuel build up and that the relationship between biomass burning and vegetation tends to vary in space, being highly regionalized (Blarquez et al., 2015). The relationship between fire and vegetation was the strongest in area where dominance shifts between fuel types occurred (e.g. the boreal mixedwood forests, Higuer et al., 2009, Girardin et al., 2013, Blarquez et al., 2015). Consequently, the relationship between biomass burning and vegetation biomass can not here be assessed beyond any doubt because of the inclusion of multiple vegetation zones (Blarquez et al., 2015).

**4.2. MAT accuracy and performance**

In previous work, the MAT has been used to quantitatively reconstruct continental-scale patterns of vegetation from remote sensing data (Williams, 2002). Here we showed that MAT technique can also perform well at reconstructing the genera biomass from pollen assemblages. Those biomass estimates are likely prone to certain source of bias and inaccuracies discussed below, among which is the propagation of the errors related to the modern datasets (Beaudoin et al., 2014, Whitmore et al., 2005) and transfer functions, that fall into the category of ‘known unknowns’ in the sense of Jackson (2012). A relationship between pollen production and dispersal and the tree biomass should be nonetheless quantifiable. Some models linking pollen deposition and source areas have been previously used to reconstruct plant communities from pollen assemblages (Sugita, 1994). Models that aim to reconstruct plant cover at the landscape scale from large lakes have been developed, mainly in Europe and Scandinavia in particular (Gaillard et al., 2010). Yet those models are not parameterized for boreal ecosystems of North America. Here we simply accommodated for differential productivity and dispersal of the different taxa by using different sized window for assessing current vegetation based on empirical relationships. However pollen productivity is controlled by climate (Hicks, 1999) and in some cases increasing tree density can reduce pollen production (Broström et al., 2008). But, in the absence of strictly independent proxies for climate, vegetation biomass and pollen in the past it is difficult to assess the relative effect of climate on pollen production. Here, because we used an analogue approach based on the principle of uniformitarianism, we must therefore assume that the relationships observed in the present hold in the past, such as the relationship between a pollen assemblage and the plant biomass. This constitutes a limitation of our approach, especially because carbon allocation strategies may have changed in response to climatic and/or disturbance events, thus modifying the relationship between pollen production and plant biomass. This relationship is thus not necessarily constant in time, however MAT involved at comparing the whole pollen assemblage typology to biomass and not that of a single taxon, likely smoothing this production bias. We also guarded against reconstructing unrealistic biomass values by carefully screening for no-analogues using the ROC curve approach (Gavin et al., 2003). These no-analogue samples were numerous during the late-glacial (14000-10000 BP, Figure 4) a period where no-analog climates and plant communities are known to be widespread (Williams and Jackson, 2007).

We tried to limit taphonomic issues by using only lacustrine environments from the modern dataset used to produce the transfer functions (Whitmore et al., 2005). Nevertheless, we did not consider the
age of the samples in the modern database, and in few occasions recent land use modifications may have affected the relationship between old (>50 yr, c. 1960) modern pollen samples and current remote sensed vegetation. But the large search radius for current biomass (>15 km) should limit uncertainties related to land cover changes that occurred during the last 50 years.

The biomass gradient used for the MAT is not even, with slightly more samples in the lower part of the biomass gradient (Figure 3), which may result in biased RMSEP and a lower accuracy of the MAT at reconstructing high biomass values (Telford and Birks, 2011). However, we showed that the MAT has a general tendency to underestimate high biomass values, so our reconstructions should be regarded as minimal biomass estimates, and it cannot be excluded that higher biomass has occurred in the past. Because MAT is known to be sensitive to autocorrelation and to the spatial structure of the data (Guiot and De Vernal, 2011, Telford and Birks, 2009), we tested parametric Weighted Averaging transfer functions (results not shown, Simpson, 2007). Nonetheless, these functions proved to be unable to reconstruct past biomass because they tended to extrapolate biomass at values below zero (Birks et al., 2010), which is nonsensical for biomass. Contrary, MAT reconstructed values were, by definition, bounded by the modern biomass dataset range, avoiding negative biomass but limiting the reconstruction to the highest current values. That being said, question remains on the use of the MAT technique itself (Birks et al., 2010, Telford and Birks, 2009) and transfer functions based on modern samples from environments potentially impacted by human activities (Anderson, 2014). Our reconstructions indeed are inherently dependent upon the quality of the modern datasets (Beaudoin et al., 2014, Whitmore et al., 2005), and we showed that uncertainties could be high notably in areas impacted by human activities (Figure S1). Then, more studies using modern relationships and experiments that aim to constrain pollen source area in North America and to quantify pollen production bias are clearly needed.

5. Conclusion

We provided here a quantitative estimate of the tree biomass for the main six tree genera of eastern Canada ecosystems over the last 14000 years. This study complements previous quantitative estimates of Quaternary forest tree cover and biome reconstructions from pollen percentages mapping (Williams, 2002, Williams et al., 2004). Particularly, we were able to provide past biomass estimates for genera (e.g., Abies and Populus) whose pollen is generally under-represented in sediments and whose abundance is difficult to estimate at the landscape scale. Our reconstructions indicate that the region supported at least 12 Pg of tree biomass within the eastern Canadian forest ecosystems, a value that was attained 7000 years ago concomitantly with the final collapse of the Laurentide Ice Sheet and in agreement with previous estimates (Figure 6, Williams et al., 2011). The total biomass trend followed the deglaciation of the region and saturated rapidly within the region. Trees were probably able to survive in a periglacial environment at the ice margin and then expand from there, explaining why trees migrated so quickly and biomass saturated so rapidly (Feurdean et al., 2013), with notably biotic velocities that exceed climatic ones at the northern fringe of postglacial tree migration (Ordonez and Williams, 2013).

The mean total genera biomass was maximal only during the HTM, when higher than present insolation, higher CO₂ levels than during the LGM, and temperature and precipitation close to present day levels favored higher tree biomass notably for hardwood species. Interestingly, this period that corresponds to major climatic and biotic rearrangements in the region, has been shown to support higher than present biodiversity within the boreal region (including α, β and γ diversity, Blarquez et al., 2014a). This could indicate that the biodiversity and biomass drivers were similar. Although it has been shown that climatic conditions likely explain both (this study, Shuman et al., 2005), the limited tree competition immediately after the deglaciation and the young stand ages at the landscape scale may also explain the elevated biomass accumulation (Ma et al., 2012). Contrary, after the initial ice retreat and the HTM period, the ecosystems have matured, soils have been formed, the biogeochemical cycles have tended.
to become more complex, and the interactions between species were more strongly consolidated, likely restricting both tree biomass production and diversity (Blarquez et al., 2014a). Future conditions would probably not have analogue in the past, and the conditions that will prevail in the future (human impact, CO$_2$ concentration, climate, disturbances) may be well beyond past conditions. But, the results of this study can be used to estimate past carbon stocks and fluxes or to infer the mechanisms linking biomass to climate and the range within which tree communities were resilient to environmental changes. Finally quantified past biomass estimates should be useful for assessing the role of land cover on regional climate forcing (Strandberg et al., 2014) or to test and validate hypotheses using dynamical vegetation modelling.

6. Acknowledgements

Data were obtained from the Neotoma Paleoecology Database (http://www.neotomadb.org), North American Surface Sample Dataset (http://www.geography.wisc.edu/faculty/williams/lab/Downloads.html), and Global Charcoal Database (http://paleofire.org), and the work of the data contributors and the community is gratefully acknowledged. We thank two anonymous reviewers for comments that greatly improved the quality of this manuscript.

7. Biosketches

Olivier Blarquez is assistant professor at the geography department of the University of Montreal. His work focuses on understanding the interactions between natural disturbances, vegetation and climate on the long-term for cold forest ecosystems (Alpine and boreal).

Julie Aleman is a postdoctoral research associate at the department of Ecology and Evolutionary Biology of Yale University. She is an ecologist interested in understanding vegetation dynamics at large spatial and temporal scales.

References


Published by NRC Research Press


Published by NRC Research Press


8. Figures and tables

<table>
<thead>
<tr>
<th></th>
<th>Abies</th>
<th>Acer</th>
<th>Betula</th>
<th>Picea</th>
<th>Pinus</th>
<th>Populus</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.497</td>
<td>0.68</td>
<td>0.923</td>
<td>1.12</td>
<td>0.714</td>
<td>0.523</td>
<td>0.878</td>
</tr>
<tr>
<td>Biomass burning</td>
<td>0.878</td>
<td>0.68</td>
<td>0.923</td>
<td>1.12</td>
<td>0.714</td>
<td>0.523</td>
<td>0.878</td>
</tr>
<tr>
<td>JJA Insolation</td>
<td>0.743</td>
<td>0.678</td>
<td>0.928</td>
<td>1.12</td>
<td>0.714</td>
<td>0.523</td>
<td>0.878</td>
</tr>
<tr>
<td>Ice area</td>
<td>0.195</td>
<td>0.194</td>
<td></td>
<td>-1.282</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta^{18}O ) NGRIP</td>
<td>-0.011</td>
<td>0.409</td>
<td>0.142</td>
<td>-2.988</td>
<td>0.35</td>
<td>0.343</td>
<td></td>
</tr>
<tr>
<td>JJA temperature</td>
<td>0.302</td>
<td></td>
<td></td>
<td>1.008</td>
<td></td>
<td></td>
<td>0.224</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>-0.131</td>
<td>-0.054</td>
<td>-0.294</td>
<td>0.935</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( CO_2 )</td>
<td>1.104</td>
<td>0.996</td>
<td>1.132</td>
<td>0.645</td>
<td></td>
<td></td>
<td>0.704</td>
</tr>
</tbody>
</table>

Table 1. Summary of the LASSO analysis to model mean tree biomass for the main genus and the mean total biomass against environmental variables. Regression coefficients for the variables that have been selected by the LASSO regularization method used for model selection are shown. “.” indicates variables not selected by the model. Prior to analysis variable have been rescaled using Z-score to enable comparing the regression coefficients. JJA temperature and annual precipitation were expressed as the mean of the HadCM3 and CCSM3 model runs.
Fig. 1. Present day biomass in t.ha$^{-1}$ for the six studied genera from Beaudoin et al. (2014) and the location of modern pollen samples (grey dots) from the North American Surface Sample Dataset (Whitmore et al., 2005). The spatial resolution of the biomass raster has been reduced to 10 * 10 km compared to the original 250 * 250 m original resolution for display purpose only.
Fig. 2. Results from the procedure used to calculate the optimal radius for compiling the biomass around each sample of the modern pollen database for each genus, based on a compromise between high determination coefficient ($r^2$ blue line), $d(r^2)$ close to or equal to 0 (red lines) and small distance. The selected distances are indicated for each genus on the graphs (grey dashed lines).
Fig. 3. Predicted vs observed aboveground biomass for the six tree genus in Eastern Canada based. Determination coefficients, Root Mean Square Error of Predictions and the number of analogues used in the Modern Analogue Technique for each genus are indicated on the graphs. Red lines corresponds to linear regressions applied to predicted vs observed biomass (red dashed lines represents 95% confidence interval on linear regressions and blues lines the prediction intervals).
Fig. 4. Diagnostic plots for the Modern Analogue Technique. (a-d) Location of the records from Neotoma database used to reconstruct the past biomass at 12000, 8000, 4000 and 0 BP. (e) Total number of samples available within each 1000 year time interval from 14000 to 0 BP. (f) Percentage of samples without analogue in the modern database within each 1000 years time interval i.e. samples excluded by the Receiver Operating Characteristic curve approach (Gavin et al., 2003).
Fig. 5. Biomass maps for the six genus and total biomass for the last 13 000 years and for present day (kNN MODIS, Beaudoin et al., 2014). The millennium period centered on 14 000 BP has not been mapped due to the low number of sites available for the mapping procedure. Color ranges have been defined by binning biomass values within five equal bins representing each 20% of the data (except for Acer were quartiles, i.e. 25% have been used for clarity). For an overview of all periods and to download the data (georeferenced raster tif files) refer to the on-line interactive application available at http://blarquez.com/maps.
Fig. 5. Continued.
Fig. 6. Temporal trend of mean genera biomass compared to climatic and physical variables. (a) Mean tree genera biomass temporal trends obtained by fitting a LOWESS (with a 200 years half window width) on reconstructed biomass from all sites. (b) Total biomass in 1000 years windows obtained by calculating the total biomass for each genera in each 50 × 50 km pixel and then summing all pixel values within the studied territory. Total biomass values were expressed as petagrams (Pg) for convenience. (c) Biomass burning trend obtained by the analysis of charcoal series contained in the Global Charcoal Database. (d) Laurentide ice sheet trend from Dyke (2004). (e) NGRIP δ18O record (grey line) and trend evaluated using 1000 years windows averages (plain red line). (f) Epica Dome C CO2 from Monnin et al. (2004). Summer (June, July, August) temperatures (g) and annual precipitation (h) anomalies from the HadCM3 and CCSM3 GCM experiments. (i) Summer (JJA) insolation anomalies at 45°N.