



APPLYING PLANT FUNCTIONAL TYPES TO CONSTRUCT BIOME MAPS FROM EASTERN NORTH AMERICAN POLLEN DATA: COMPARISONS WITH MODEL RESULTS

JOHN W. WILLIAMS, ROBERT L. SUMMERS and THOMPSON WEBB III

Department of Geological Sciences, Brown University, Providence, RI 02912-1846, USA

(e-mail: Jww@brown.edu)

Abstract—Global biome models like BIOME1 convert climate-model simulations of past climates into biome distributions and thus facilitate comparison of both climate and biome model results with biomes estimate from paleoecological data. We adapted a biomization method, recently developed for European pollen data, for use with pollen data in eastern North America and then compared its estimated biomes with those derived from applying BIOME1 to the climate simulations from the NCAR CCM1 (National Center for Atmospheric Research Community Climate Model, Version 1) for 6000 years ago (6 ka). We first tested the biomization method by seeing how well the biomes inferred from modern pollen data match observed biomes. We found that modifications to the method were necessary in part to account for physiological differences between North American and European taxa, and in part to cope with our choice of using just 23 major pollen taxa. Our modifications significantly improved the match between observed modern biomes and pollen-derived biomes, as measured by the kappa statistic. We tested our tuning of the biomization method by matching its inferred 6 ka biomes to biomes estimated from pollen data using the modern analog technique. The degree of agreement at 6 ka is close to that for today, showing that (1) the biomization method and modern analog technique, when applied to the same pollen data, produce consistent results, and (2) the modifications made to the biomization method are robust back to 6 ka. We then used the results of the biomization method to test the biome maps simulated by BIOME1, which derives biome distributions from observed climate values for today and from the climatic simulations of the CCM1 for 6 ka. Only a fair agreement is seen, and significant offsets exist in the placement of biomes by BIOME1. For today BIOME1 simulates the boundary between the temperate deciduous and cool mixed forests to be too far south and the steppe-forest boundary to be too far west. These model biases are also evident in the simulations at 6 ka despite the fact that CCM1 simulates warmer than present temperatures in the central United States. To the north, however, BIOME1 correctly simulates the cool mixed forest and taiga boundary at 6 ka as more north-westward than at present. © 1998 Elsevier Science Ltd. All rights reserved



INTRODUCTION

The development of global biome models represents a new stage in earth system modeling. By translating simulated climate variables into maps of biomes, these models explicitly link vegetation and climate patterns together, enabling the climate simulations of general circulation models (GCMs) to be evaluated by comparison to observed or inferred vegetation distributions. Just as GCMs extend our understanding of the climate system into areas and times with few direct measurements of climate, biome models can estimate vegetation for past times and regions containing few paleovegetation data and can also predict the vegetational response to increased atmospheric CO₂ conditions (VEMAP members, 1995). The fundamental units of most biome models are plant functional types (PFTs), which group plant taxa according to their ecological similarities rather than evolutionary heritage (Prentice *et al.*, 1992). Biomes, which are plant assemblages of large enough scale that

their extent is primarily controlled by climate (Clements and Shelford, 1939), are then defined in biome models as assemblages of PFTs. BIOME1, the biome model evaluated in this paper, predicts vegetation distribution from four bioclimatic variables: growing degree days, mean temperature of the coldest month, mean temperature of the warmest month, and soil moisture (Prentice *et al.*, 1992). Each PFT is assigned a set of tolerances, which frequently overlap among PFTs, and biomes are defined as the various possible combinations of PFTs.

By presenting climate simulations from GCMs in terms of biome distributions, biome models show how the summed effect of climate variables changes through time, and allow the simulations of both biome models and GCMs to be evaluated by comparing the simulation to biomes observed or derived from fossil data. Because biome models are generally tuned to modern vegetation–climate relations, the strongest test of a biome model is to compare its ‘postdicted’ vegetation

distributions with biome maps inferred from pollen or macrofossil data. Many workers have mapped biomes for past time periods (Frenzel, 1973; Frenzel *et al.*, 1992; COHMAP members, 1988; Guetter and Kutzbach, 1990; Adams *et al.*, 1990; Delcourt and Delcourt, 1981), but only in the past several years have quantitative methods for inferring biomes from pollen data been developed (Overpeck *et al.*, 1992), of which the biomization method (Prentice *et al.*, 1996) is the most recent.

No matter the method, any vegetation reconstruction based upon pollen evidence must deal with the complex and biased nature of the signal. The major advantage of pollen data is that they provide a quantitative record that can be treated statistically, but unfortunately pollen abundances cannot be translated directly into plant abundances due to the inequalities in pollen production, dispersal, and preservation among plants (Faegri *et al.*, 1989; Prentice, 1988). Plant taxa are usually over- or under-represented in the pollen record, so similar pollen abundances of two taxa may represent very different plant abundances. Furthermore, a pollen collection site, typically a lake or bog, collects pollen from a surrounding area with a radius of approximately 100 km (Bradshaw and Webb, 1985). This fact means that local and regional signals can be confounded, although this problem can be lessened by choosing to study taxa that are known to spread their pollen either locally or widely. More importantly, lakes

act to mix pollen from different distances and thus smooth out the local gradients in vegetation composition (Jacobson and Bradshaw, 1981). This mixing aids the effort to reconstruct biomes, because biomes are features of the regional vegetation in which local patches are only a source of variation, which can be minimized by appropriate sampling strategies (Webb, 1993).

Prentice *et al.* (1996) developed the biomization method to infer biomes from pollen data, and sought to overcome some of the complexity inherent in pollen data by grouping taxa at the level of PFTs rather than individual species. They also set threshold values to remove the noise from insignificant pollen abundances, and used square roots to downweight the most over-represented pollen taxa. The method is designed to be flexible for global use by using broadly defined PFTs that allow it to de-emphasize regional differences in plant assemblages and to focus on the broad-scale nature of biomes. Furthermore, the biomization method is intended to be compatible to BIOME1: they share a similar logic in defining biomes as assemblages of plant functional types, and the number and types of biomes constructed can be adjusted within the biomization method to match the classification scheme used by BIOME1. Prentice *et al.* (1996) successfully applied the method to European data, and we have extended the method to pollen data from eastern North America, as part of an international project, BIOME 6000,

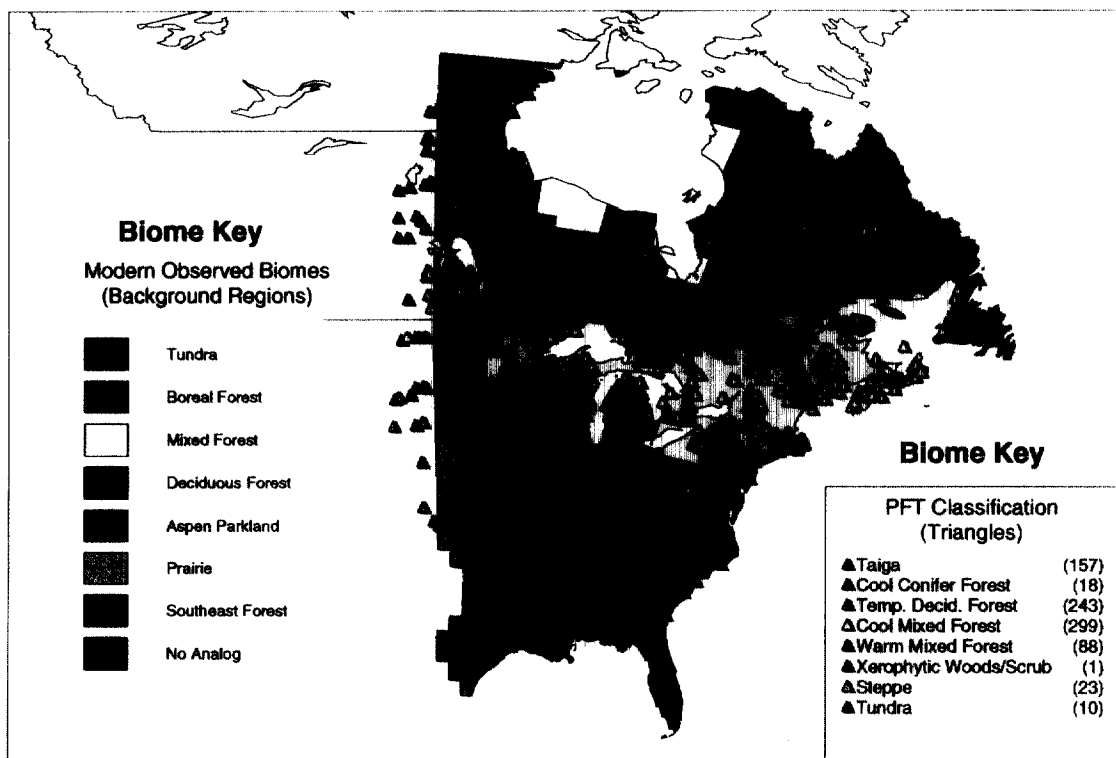


FIG. 1. Map showing pollen sites and their biomes as inferred by the first iteration of the biomization method (triangles) overlain upon a map of observed and potential biomes compiled by Overpeck *et al.* (1992) (background regions). The key for each biome scheme is shown separately, but the color codes for the two schemes are nearly the same. A triangle and background of the same color thus indicates agreement between the biomization method and modern observed vegetation, and a triangle differently colored from its background indicates a discrepancy.

TABLE 2. The relationship between the biome classifications of Overpeck *et al.* (1992) and this paper.

Overpeck <i>et al.</i> (1992)	Prentice <i>et al.</i> (1992), this paper
Tundra	Tundra
Forest Tundra	Taiga
Boreal Forest	Taiga
	Cool Conifer Forest
Mixed Forest	Cool Mixed Forest
Deciduous Forest	Temperate Deciduous Forest
Prairie	Steppe
Southeast Forest	Warm Mixed Forest
Aspen Parkland	
No analog*	

*'No analog' is used by Overpeck *et al.*, to mark vegetation assemblages that are not present today.

turn are based upon the modern biomes described by Olson *et al.* (1984). This consistency allows a direct comparison between our results and the vegetation simulations of BIOME1 produced from CCM1 (Kutzbach *et al.*, 1998). Overpeck *et al.* (1992) define their biomes slightly differently, but most of their biomes have direct counterparts in this paper (Table 2). The boreal forest and forest tundra biomes of Overpeck *et al.* (1992) are in this paper combined under the name boreal forest, which is equivalent to the taiga biome of Olson *et al.* (1984). The cool conifer biome used in this paper does not have a direct counterpart in Overpeck *et al.* (1992), but marks a transitional zone between their cool mixed forest and boreal forest biomes. Conversely, we do not consider aspen parkland, a minor biome in ENA composed of open forests of *Populus*.

We use previously generated biome maps for today and 6 ka (Overpeck *et al.*, 1992) to assess our results. The map of modern vegetation was compiled from previous maps of observed and potential vegetation (Bernabo and Webb, 1977; Rowe, 1972; Kuchler, 1964). To infer biomes for 6 ka, Overpeck *et al.* (1992) used the modern analog technique (MAT), first developed by Overpeck *et al.* (1985). Their method has five steps: (1) a database of modern pollen samples is obtained for the region of interest; (2) each modern pollen sample in the database is linked to the plant assemblage surrounding the collection site; (3) each fossil pollen sample *i* is compared to all modern pollen samples using the squared-chord distance (SCD) measure of dissimilarity (Overpeck *et al.*, 1985); (4) all modern pollen samples with a $SCD \leq 0.15$ define a subset of likely modern analogs to sample *i*; (5) the biome most represented by the pollen samples in this subset is assigned to sample *i* (R. Webb, pers. comm.). A fossil pollen sample is considered to have no modern analog if no modern samples are sufficiently similar (all $SCD > 0.15$), but this situation rarely occurred for pollen samples from 6 ka (Overpeck *et al.*, 1992).

Models Tested and the Kappa Statistic

The maps produced by BIOME1 and shown in this paper are based upon the climate simulations of

CCM1 (Kutzbach *et al.*, 1998). CCM1 improves upon CCM0 by incorporating interactive treatments of soil moisture, changes in snow cover, sea ice, and sea surface temperatures, thus having a reduced reliance upon prescribed boundary conditions (Kutzbach *et al.*, 1998). The boundary conditions for CCM1 simulations are updated and include atmospheric CO₂ levels from ice core records (Oeschger *et al.*, 1983; Lorius *et al.*, 1984; Barnola *et al.*, 1987), revised estimates of the height of northern hemisphere ice sheets from Peltier (1994), and the use of calendar years to allow changes in orbital forcing parameters (Bard *et al.*, 1990; Stuiver and Reimer, 1993) to mesh better with other boundary conditions.

BIOME1 is a mechanistic model that predicts the location of plant types based a set of bioclimatic variables carefully chosen to reflect the mechanisms of plant-climate interactions, thus reducing the need for empirical correspondences between the distribution of plant types and modern climate (Prentice *et al.*, 1992). The basic vegetation units of BIOME1 are PFTs, which categorize plant taxa according to life form and climatic constraints. Life form categories include herb, shrub, conifer, broad-leaved trees, and deciduous vs. evergreen trees. These categories are further grouped according to their tolerances for four key bioclimatic parameters: (1) a cold tolerance, or the minimum temperature that the taxon can endure, (2) a chilling requirement, which reflects the need for woody boreal taxa to have a sufficiently cold winter in order to achieve an adequate springtime budburst, (3) a requirement for a sufficiently warm growing season, and (4) a moisture requirement (Prentice *et al.*, 1992). Climatic constraints for plant taxa ideally are used only if the limiting physiological mechanism is well understood, or if the constraint has been determined by controlled experiments. Because PFTs are defined in terms of climatic constraints, PFT ranges may be simulated from climatic variables, either observed or simulated. The distribution of biomes (defined as assemblages of PFTs) may in turn determined from PFT ranges.

Confusion matrices, which have been used extensively in remote sensing of vegetation classification (e.g. Ranson and Sun, 1994; Ranson *et al.*, 1995; Nezry *et al.*, 1993; Arai, 1993), can be used to tabulate the degree of agreement between biomes derived from the biomization method and biomes observed or calculated by other methods (Table 3). Each row of a confusion matrix corresponds to a biome region of an underlying map, each column corresponds to a biome category of the biomization method, and each entry (a_{ij}) represents the number of sites in the MAT or BIOME1 biome region *i* classified by the biomization method to biome *j*. Thus, sites that show an agreement between the biomization method and its target fall along the main diagonal, and sites that show a disagreement are counted in off-diagonal entries. The overall similarity between two biome classifications is summarized by the

TABLE 3. Confusion matrices recording the match between the biomization method and modern observations, the modern analog technique (MAT), and BIOME1

Modern Potential Biomes (Overpeck <i>et al.</i> , 1992)	PFT method									Total
	Tundra	Taiga	Cool Conifer Forest	Cool Mixed Forest	Temperate Deciduous Forest	Warm Mixed Forest	Steppe	Aspen Parkland	No analog	
(a) Modern, 1st iteration										
Tundra	9	17	0	0	0	0	0	—	—	26
Boreal Forest	0	122	16	72	1	0	0	—	—	211
Cool Conifer Forest	—	—	—	—	—	—	—	—	—	—
Mixed Forest	0	0	2	185	56	0	0	—	—	243
Deciduous Forest	0	0	0	24	152	20	1	—	—	197
Southeast Forest	0	0	0	0	23	66	1	—	—	90
Prairie	1	0	0	0	8	2	7	—	—	18
Aspen Parkland	0	0	0	6	0	0	0	—	—	6
No Analog	0	0	0	0	0	0	0	—	—	0
Total	10	139	18	287	240	88	9	—	—	791
(b) Modern, 2nd iteration										
Tundra	13	13	0	0	0	0	0	—	—	26
Boreal Forest	5	117	16	73	0	0	0	—	—	211
Cool Conifer Forest	—	—	—	—	—	—	—	—	—	—
Mixed Forest	0	0	2	234	7	0	0	—	—	243
Deciduous Forest	0	0	0	59	133	4	1	—	—	197
Southeast Forest	0	0	0	6	18	64	2	—	—	90
Prairie	0	0	0	1	6	0	11	—	—	18
Aspen Parkland	0	0	0	6	0	0	0	—	—	6
No Analog	0	0	0	0	0	0	0	—	—	0
Total	18	130	18	379	164	68	14	—	—	791
(c) Modern, 3rd iteration										
Tundra	13	13	0	0	0	0	0	—	—	26
Boreal Forest	5	175	19	12	0	0	0	—	—	211
Cool Conifer Forest	—	—	—	—	—	—	—	—	—	—
Mixed Forest	0	1	38	192	12	0	0	—	—	243
Deciduous Forest	0	0	0	38	151	7	1	—	—	197
Southeast Forest	0	0	0	1	18	70	2	—	—	91
Prairie	0	0	0	0	6	0	12	—	—	18
Aspen Parkland	0	0	1	4	0	0	1	—	—	6
No Analog	0	0	0	0	0	0	0	—	—	0
Total	18	189	58	247	187	77	16	—	—	791
(d) Modern, 'Optimal'										
Tundra	5	5	0	0	0	0	0	—	—	10
Boreal Forest	1	151	10	6	0	0	0	—	—	168
Cool Conifer Forest	—	—	—	—	—	—	—	—	—	—
Mixed Forest	0	1	31	147	3	0	0	—	—	182
Deciduous Forest	0	0	0	15	113	3	1	—	—	132
Southeast Forest	0	0	0	0	4	59	1	—	—	64
Prairie	0	0	0	0	5	0	11	—	—	16
Aspen Parkland	0	0	0	0	0	0	0	—	—	0
No Analog	0	0	0	0	0	0	0	—	—	0
Total	6	157	41	168	125	62	13	—	—	572

kappa statistic (Prentice *et al.*, 1992; Monserud and Leemans, 1992; Cohen, 1960). The kappa statistic is calculated as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where $p_o = \sum_{i=1}^m a_{ii}/n$ is the observed proportion of agreement (a_{ii} are the entries lying on the main diagonal; n is the total number of sites considered) and $p_e = \sum_{i=1}^m r_i c_i / n^2$ is the proportion of agreement expected if all agreements occurred by chance alone (r_i and c_i are row and column totals).

TABLE 3. (Continued)

MAT (Overpeck <i>et al.</i> , 1992)	PFT method									Total
	Tundra	Taiga	Cool Conifer Forest	Cool Mixed Forest	Temperate Deciduous Forest	Warm Mixed Forest	Steppe	Aspen Parkland	No analog	
(e) 6 ka, 3rd iteration										
Tundra	7	0	0	0	0	0	0	—	—	7
Boreal Forest	4	27	5	0	0	0	0	—	—	36
Cool Conifer Forest	—	—	—	—	—	—	—	—	—	—
Mixed Forest	0	4	9	111	4	3	0	—	—	131
Deciduous Forest	0	0	0	19	37	1	1	—	—	58
Southeast Forest	0	0	0	0	2	9	0	—	—	11
Prairie	0	0	0	0	2	0	10	—	—	12
Aspen Parkland	0	0	0	0	0	0	1	—	—	1
No Analog	0	0	0	4	0	0	0	—	—	4
Total	11	31	14	134	45	13	12	—	—	260
(f) 6 ka, 'Optimal'										
Tundra	5	0	0	0	0	0	0	—	—	5
Boreal Forest	2	25	3	0	0	0	0	—	—	30
Cool Conifer Forest	—	—	—	—	—	—	—	—	—	—
Mixed Forest	0	4	9	95	0	1	0	—	—	109
Deciduous Forest	0	0	0	7	27	0	0	—	—	34
Southeast Forest	0	0	0	0	1	9	0	—	—	10
Prairie	0	0	0	0	1	0	8	—	—	9
Aspen Parkland	0	0	0	0	0	0	1	—	—	1
No Analog	0	0	0	1	0	0	0	—	—	1
Total	7	29	12	103	29	10	9	—	—	199
PFT method										
BIOME1 (Prentice <i>et al.</i> , 1992)	Tundra	Taiga	PFT method					Steppe	Total	
			Cool Conifer Forest	Cold Mixed/Cold Deciduous	Cool Mixed Forest	Temperate Deciduous Forest	Warm Mixed Forest			
(g) Modern, compared with BIOME1										
Tundra	17	31	0	—	0	0	0	0	0	48
Taiga	1	122	1	—	2	0	0	0	0	126
Cool Conifer Forest	0	40	14	—	15	1	0	0	0	70
Cold Mixed/Cold Deciduous	0	3	1	—	3	1	0	4	12	
Cool Mixed Forest	0	4	41	—	202	108	0	2	357	
Temperate Deciduous Forest	0	0	0	—	11	62	16	2	91	
Warm Mixed Forest	0	0	0	—	1	11	56	2	70	
Steppe	0	0	1	—	0	1	0	23	25	
Total	18	200	58	—	234	184	72	33	799	
(h) 6 ka, compared with BIOME1										
Tundra	3	0	0	—	0	0	0	0	0	3
Taiga	5	26	1	—	6	2	0	1	41	
Cool Conifer Forest	1	7	5	—	15	5	0	1	34	
Cold Mixed/Cold Deciduous	0	0	0	—	0	3	0	2	5	
Cool Mixed Forest	0	2	8	—	95	23	4	5	137	
Temperate Deciduous Forest	0	0	0	—	14	8	1	0	23	
Warm Mixed Forest	0	0	0	—	0	2	6	0	8	
Steppe	0	0	0	—	0	0	0	5	5	
Total	9	35	14	—	130	43	11	14	256	

Monserud (1990) proposed that the agreement between two vegetation classification schemes be rated as follows: <0.4 is poor, 0.4–0.55 fair, 0.55–0.7 good, 0.7–0.85 very good, and >0.85 excellent. Individual kappa scores may be calculated for biomes according to the following equation:

$$\kappa_i = \frac{\frac{a_{ii}}{n} - \frac{r_i c_i}{n^2}}{\frac{r_i + c_i}{2n} - \frac{r_i c_i}{n^2}} \quad (\text{Monserud and Leemans, 1992}).$$

The kappa statistic requires that both classification methods use the same categories. This requirement is fulfilled when BIOME1 and the biomization method are compared, but as mentioned above the MAT and biomization method use slightly different definitions. To calculate the kappa scores for this section, we did not count any sites that the biomization method classified as cool conifer forest, and matched the biomes of Overpeck *et al.* (1992) to their closest PFT counterpart (Table 2).

A simple alternative to the kappa statistic is to calculate the frequency of ‘hits’ (correctly inferred sites) in each biome (Table 7) and to calculate a weighted average of these frequencies to obtain an overall statistic of agreement. Although this method is simpler, its

disadvantage is that it only measures along-row agreements and does not take into account disagreements along columns. For example, in Table 3b, the seven sites in the MAT-inferred tundra are all classified by the biomization method as tundra, giving tundra a perfect score of 1.0 in the corresponding hit frequency table (Table 6). This ignores, however, four sites classified by the biomization method as tundra in the MAT-inferred boreal forest (Table 3b). The kappa statistic, which accounts for disagreements in both columns and rows, captures the differences in tundra classifications indicated by these sites ($\kappa_T = 0.82$). Both the kappa statistic and hit frequencies are presented for purposes of comparison, but the overall agreement is better captured by the kappa statistic than it is by the simpler hit frequencies.

THE BIOMIZATION METHOD

The biomization method (Prentice *et al.*, 1996) associates biomes with pollen samples (modern or fossil) in five steps by (1) assigning pollen taxa to PFTs, (2) assigning PFTs to biomes, (3) combining the results of steps (1) and (2) to assign pollen taxa to biomes, (4) calculating affinity scores between each pollen sample and biome, and (5) assigning each pollen sample to the

TABLE 4. Assignments of plant functional types (PFTs) to biomes for eastern North America

Plant functional types	Biomes										
	Taiga	Cool Conifer Forest	Cool Mixed Forest	Temperate Deciduous Forest	Warm Mixed Forest	Tundra	Steppe	Cold Deciduous Forest	Cold Mixed Forest	Xerophytic Woods/ Scrub	Desert
Boreal Evergreen Conifer	X	X	X								
Cool Temperate Conifer		X	X	(X)				X			
Eurythermic Conifer	X	X	X	(X)	X		X	X	X		
Boreal Summergreen	X	X	X	X			X	X			
Cool Temperate Summergreen		X	X	X	X			X			
Temperate Summergreen			X	X	X						
Warm Temperate Summergreen				X	X						
Warm Temp. Broad-Lvd. Evergrn.					X					X	
Arctic Alpine Shrub						X					
Sedge						X					
Steppe Forbs							X				
Desert Forbs											X
Intermediate Temperate Conifer			X	X							
Cool Temperate Evergreen				X	X						
Warm Temp. Schlerophyll Shrub										X	
Grass						X	X				X
Health	X	X	X	X	X	X		X	X		

(X) indicates that the PFT originally belonged to that biome but was removed by us.

X indicates that the PFT originally did not belong to that biome but was added by us.

Biomes and PFTs not in boldface were not considered in this paper; they are included for completeness.

biome with the highest affinity for that sample and resolving ties if necessary. The biomization method as developed by Prentice *et al.* (1996) is described below, and our modifications are set forth in a later section.

The biomization method requires that pollen taxa explicitly be assigned to PFTs (Table 1). A pollen taxon in theory should belong to only one plant functional type, but this is not always the case because pollen taxa are usually identifiable only to genus or family level, and species from the same genus can have different life forms and/or climatic tolerances. For example, *Betula papyrifera* and *B. alleghaniensis* are, respectively, boreal and cool-temperate summergreen trees found across northeastern North America, but give way in the Canadian boreal forest to *Betula nana*, which grows as an Arctic alpine shrub. Similarly, *Pinus* in eastern North America today consists of two separate populations: northern pines (e.g. *P. strobus*, *P. banksiana*, *P. resinosa*) and southern pines (e.g. *P. taeda*, *P. palustris*, *P. elliotii*).

Biomes are defined as assemblages of PFTs (Table 4) (Prentice *et al.*, 1992, 1996). Since all PFTs have inherent climate spaces, which frequently overlap, any given region may contain several PFTs, and biomes are simply the labels given to the various possible PFT assemblages. Naturally, a PFT may belong to more than one biome. Our initial PFT–biome assignments followed those of Prentice *et al.* (1996), but we later changed several assignments to match observed PFT–biome relationships in ENA today (see below).

Once taxa are assigned to PFTs and PFTs to biomes, taxa may then be matched to biomes using PFT's as an intermediary, i.e. if taxon *a* belongs to PFT *b*, and PFT *b* belongs to biome *c*, then taxon *a* is considered to belong to biome *c*. Taxa were matched to biomes by multiplying the taxon \times PFT and PFT \times biome matrices to produce a matrix with 23 rows (taxa) and 7 columns (biomes) (see Table 5).

Affinity scores measure the suitability of biomes for a given pollen sample and are calculated for all potential biomes at each site according to the following equation:

$$A_{lk} = \sum_{i=1}^n \delta_{ik}(p_{il} - \theta_i)^{0.5}$$

where A_{lk} is the affinity score of biome *k* for sample *l*, p_{il} the pollen percentage of taxon *i* in sample *l*, δ_{ik} , obtained from the taxon \times biome matrix, reflects whether taxon *i* is in biome *k* ($\delta_{ik} = 0$ or 1), n is the number of taxa (in this paper, $n = 23$), and θ_i a threshold value, initially set for all taxa at 0.5%.

Affinity scores thus consist of a sum of square roots of pollen percentages, less a threshold value. Taking square roots downweights the more abundant taxa, while setting a threshold value reduces noise arising from low amounts of pollen. So that only pollen

taxa characteristic of a biome will 'vote' for that biome, each pollen percentage is multiplied by δ_{ik} , which is equal to 1 if taxon *i* is present in biome *k*, 0 otherwise. Because several biomes contain similar groups of taxa, the biomization method is at times overly dependent upon the presence or absence of certain taxa, e.g. *Betula* is critical for choosing between the temperate deciduous and warm mixed forest biomes (Table 5).

Once affinity scores have been calculated for all biomes, the biome with the highest affinity score is assigned to the sample. However, ties in affinity scores often result when one biome is a subset of another (e.g. taiga is a subset of cool conifer forest which is a subset of cool mixed forest; see Tables 4 and 5), and we found that, due to our relatively short taxon list, biomes were similar enough that ties often occurred. Prentice *et al.* (1996) established a tie-break ranking of biomes to favor biomes that were subsets of other biomes. We developed an alternative algorithm, based upon the average abundance of taxa for a given time interval, that also accounts for ties when biomes are not subsets (see below).

INITIAL MODIFICATIONS AND RESULTS

Modifications

Initially we used the taxon–PFT–biome assignments determined by Prentice *et al.* (1996). However, four genera not present in Europe also had to be assigned to PFTs. *Carya* and *Celtis*, which occur throughout modern ENA forests mainly south of 45° latitude, were assigned to the temperate (broad-leaved) summergreen PFT. *Liquidambar*, highly susceptible to frost, was assigned to the warm temperate (broad-leaved) summergreen PFT (Fowells, 1975). *Tsuga* is restricted to the northern mixed forest and the southern fringe of the boreal forest and so was classified as a cool temperate conifer. Four biomes, cold deciduous forest, cold mixed forest, xerophytic woods/scrub, and desert, when reconstructed by the biomization method, had no significant presence in ENA. Experiments showed that a few pollen samples were assigned to these biomes, but they showed no coherent distribution and effectively added a source of noise to the biome maps. These biomes were included in our working runs, but were dropped when making the final runs, and the pollen samples were reassigned to the next closest biome.

Preliminary attempts to apply the biomization method to ENA pollen data revealed that in one out of three modern samples the affinity scores for two biomes were tied. Ties occur primarily when the taxa of one biome are a subset of the taxa of another (e.g. taiga, cool conifer forest, and cool mixed forest), but can also occur when two biomes have highly similar taxon lists but are not subsets. To resolve ties, Prentice *et al.* (1996) ranked biomes such that subset biomes were always favored over the more inclusive biomes, but did

TABLE 6. Kappa scores for the match between classification schemes overall and for individual biomes

	MODERN				6 KA			
	Modern Potential Biomes vs.				BIOME1 vs.	MAT vs.		BIOME1 vs.
	1st iteration (Fig. 1)	2nd iteration (Fig. 2)	3rd iteration (Fig. 3)	'Optimal'	3rd iteration (Fig. 5)	3rd iteration (Fig. 4)	'Optimal'	3rd iteration (Fig. 6)
Overall score	0.60	0.66	0.79	0.89	0.53	0.75	0.87	0.39
Standard deviation	0.02	0.02	0.02	0.02	0.02	0.04	0.03	0.05
Tundra	0.49	0.58	0.58	0.62	0.50	0.77	0.83	0.49
Taiga	0.66	0.65	0.89	0.94	0.70	0.85	0.87	0.63
Cool Conifer Forest	—	—	—	—	0.15	—	—	0.14
Cool Mixed Forest	0.56	0.62	0.79	0.89	0.51	0.75	0.87	0.38
Temp. Deciduous Forest	0.58	0.66	0.71	0.84	0.35	0.64	0.83	0.16
Warm Mixed Forest	0.71	0.79	0.81	0.93	0.77	0.74	0.89	0.62
Steppe	0.51	0.68	0.72	0.75	0.85	0.86	0.94	0.58

not provide a strict logic for ties between biomes that are not subsets. We adopted an alternative procedure that calculated the average abundance in ENA of a pollen taxon for each time slice, and ranked biomes by the sum of the average abundances of their constituent pollen taxa. When ties occurred the biome with the smaller sum was chosen. This procedure has the same effect as the Prentice *et al.* tie-break when one biome is a subset of another, but in other cases will give preference to a biome with relatively rare taxa (their presence in the pollen record is more informative) over a biome with more common pollen types. Note that because average pollen abundances are calculated for each time slice, the tie-break ranking of biomes changes somewhat over time.

Results

When applied to modern pollen data in eastern North America, the biomization method yielded a broad agreement with observed biomes, but also contained many mismatches (Fig. 1) and had an overall kappa score of $\kappa = 0.60$, suggesting a good but far from excellent agreement. Inspection of the kappa scores for individual biomes (Table 6) and the map overlay (Fig. 1) pointed to specific areas of disagreement: (1) cool mixed forest sites were scattered throughout the observed boreal forest, (2) cool conifer forest was not well represented by the biomization method; (3) the temperate deciduous/cool mixed forest boundary was placed too far north in Michigan, New York, and Massachusetts; (4) the biomization method did not distinguish coherently the temperate deciduous and warm mixed forest biomes; (5) the biomization method placed the steppe/forest boundary too far west; and (6) tundra, with the lowest individual kappa score ($\kappa_T = 0.49$), was not well represented by the biomization method. The map overlay (Fig. 1) and the indi-

vidual kappa scores (Table 6) at times highlight different disagreements. Tundra and steppe, on the one hand, had the lowest individual kappa scores because they contained the fewest number of pollen sites and were most affected by a few mismatches. On the other hand, the large number of cool mixed forest sites in the observed boreal forest is visually striking but was mitigated by the large number of taiga sites and is less noticeable in the kappa statistic for taiga ($\kappa_{Ta} = 0.66$).

Some mismatches were due to the biomization method inadequately distinguishing between biomes with highly similar taxon lists (e.g. cool mixed forest, temperate deciduous forest, warm mixed forest; see Table 5). In addition, a comparison of isopoll maps (Webb *et al.*, 1993) and modern observed biomes (Fig. 1) revealed that our initial use of taxon-PFT-biome assignments from Prentice *et al.* (1996) overlooked several important differences in taxonomic composition between ENA and European biomes. In particular, temperate deciduous forest in eastern North America can be distinguished from the cool and warm mixed forests by the relatively low pollen abundance for *Pinus* and other conifers (Webb *et al.*, 1993). The under-representation of steppe arose in part because, unlike in Europe, Cyperaceae pollen is moderately abundant in ENA steppe samples. Excluding Cyperaceae from steppe, therefore, weakened the ability of the biomization method to distinguish steppe vegetation. We changed several PFT-biome assignments to account for these observations (see below).

SECOND ITERATION OF MODIFICATIONS AND RESULTS

Modifications

As a first pass at improving the match of biomes inferred by the biomization method to modern

observed biomes, we altered three PFT–biome assignments: (1) the cool temperate conifer PFT, containing *Abies* and *Tsuga*, was removed from temperate deciduous forest; (2) eurythermic conifer, containing *Pinus*, also was removed from temperate deciduous forest; and (3) Cyperaceae was added to steppe.

We removed the cool temperate conifer PFT from temperate deciduous forest because isopoll maps of *Abies* and *Tsuga* (Webb *et al.*, 1993) show that the southern limit of their ranges coincides with the cool mixed forest/temperate deciduous boundary (Fig. 1). A similar examination for *Pinus* reveals that a central feature of modern temperate deciduous forest is a low abundance of *Pinus*. Not using *Pinus* to distinguish temperate deciduous forest from other forest biomes severely handicapped the biomization method, and we therefore removed the eurythermic conifer PFT from temperate deciduous forest (Table 4). This modification alone, however, was not sufficient because *Pinus* is a ubiquitous taxon and its pollen is present at some level in most samples from ENA. However, while *Pinus* pollen is present to some degree at sites in the temperate deciduous forest, isopoll maps indicate that pollen abundances >25% are reserved for the adjacent cool and warm mixed forest biomes (Webb *et al.*, 1993). Consequently, we raised the *Pinus* threshold (Θ_P) to 25% to ensure that only high abundances of *Pinus* would vote against temperate deciduous forest.

Results

A significantly improved map of modern biomes results from the inclusion of the new PFT–biome assignments (Fig. 2), and the overall kappa score increases to 0.65 (Table 6). The increase for the kappa score was relatively small because the affected biomes contain relatively few pollen sites, but all of the kappa scores for individual biomes either improved or held constant (Table 6). The addition of sedge to steppe improved the representation of steppe (κ_S rises from 0.51 to 0.68), but the steppe/forest boundary is still too far west. Removing *Pinus* from temperate deciduous forest (thus differentiating it from the cool and warm mixed forests) improved the representation of temperate deciduous forest (κ_{TD} rises from 0.58 to 0.66), making the biome much more internally cohesive and its boundaries with cool and warm mixed forest much closer to modern observed biome patterns. An unintended benefit of raising the *Pinus* threshold to 25% was the improved representation of tundra: this change negated small amounts of *Pinus* pollen that had previously led the biomization method to favor taiga over tundra. The modifications did not improve the representation of taiga, since cool mixed sites continued to be inferred too far north. The cool mixed forest biome was also overemphasized by the biomization method in the southern Appalachians and southern New England. The causes for these discrepancies are discussed below.

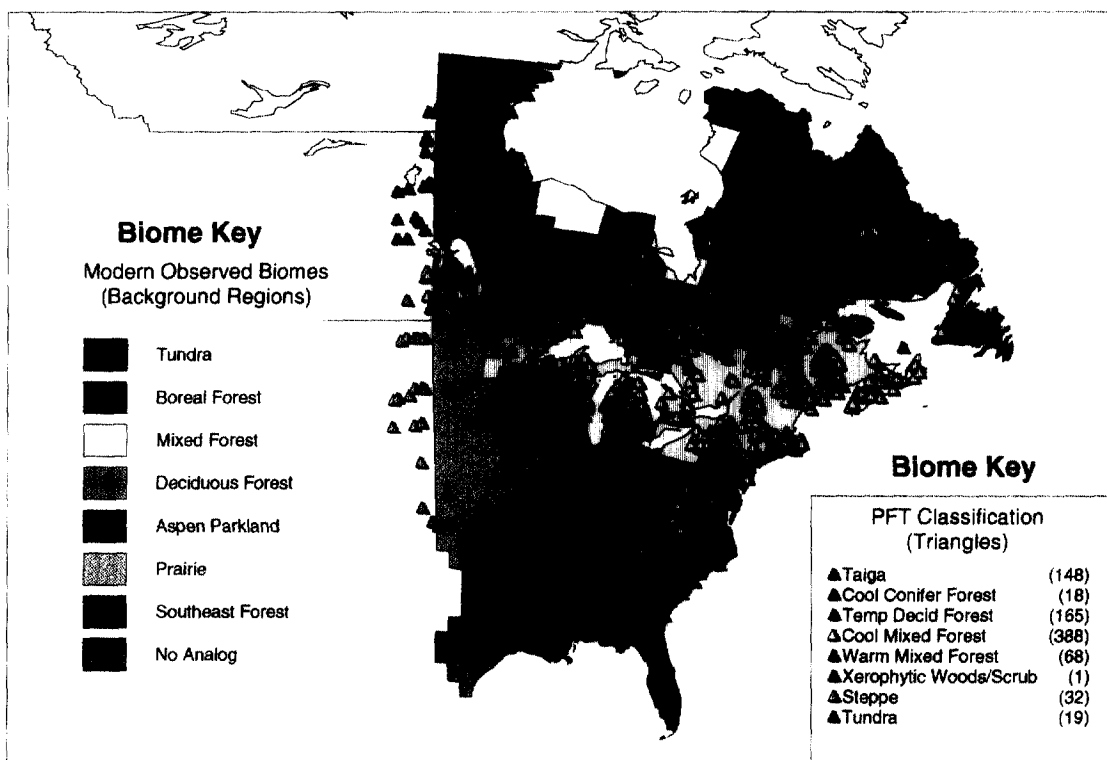


FIG. 2. The second iteration of the biomization method overlain upon modern observed and potential biomes. See the caption for Fig. 1 for a fuller description.

THIRD ITERATION OF MODIFICATIONS AND RESULTS

Modifications

The incorrect assignment of cool mixed forest sites to the observed boreal forest was mostly caused by relatively low (<1%) values of warm temperate summergreen taxa such as *Quercus*, *Ulmus*, *Fraxinus*, or *Ostrya-Carpinus*. The presence of any of these taxa in abundances greater than the threshold value of 0.5% was enough to tip the biomization method in favor of cool mixed forest over taiga (see Table 5), but the invariably low pollen percentages of these taxa suggest that the taxa are not significantly present in the boreal landscape. This suggests that the problem is mainly one of noise created by low pollen percentages present for all pollen samples, and so we raised the threshold value (Θ_i) to 1% for all taxa to decrease the impact of this source of noise.

Cool mixed forest sites continued to be found within the boreal forest, however, and were still over-represented by the biomization method in southern New England and the southern Appalachians. A survey of the pollen percentages at these sites revealed that small but greater than 1% abundances of *Quercus*, *Betula*, and *Tsuga* were responsible for this discrepancy: *Quercus* pollen favored cool mixed forest over cool conifer forest and taiga, *Tsuga* pollen favored cool mixed forest over temperate deciduous forest, and *Betula* pollen favored cool mixed forest over warm mixed forest in the southern Appalachians. Because *Quercus* and *Betula* may be over-represented in the pollen record or have significant background percentages (Bradshaw and Webb, 1985; Webb *et al.*, 1981), small pollen percentages of these taxa may be caused by a few locally prolific producers or long-distance transport rather than a significant regional presence. We provide a rough adjustment for this possibility by raising the threshold value for these pollen taxa to 2.5%. We also raised the *Tsuga* cutoff to 2.5%. This does not reflect a similar over-representation, but rather was done because the 2.5% isopoll today approximates the extent of cool mixed forest in the southern Appalachians (Webb *et al.*, 1993). This is an empirical tuning that should become unnecessary when more taxa are added to discriminate the cool and warm mixed forests.

Results

These modifications led to further improvements in the match between the PFT-derived biome classification and observed vegetation (Tables 6 and 7, Fig. 3). The overall kappa score increased to 0.79 and the overlay map revealed a high qualitative agreement between modern potential biomes and the biomes inferred by the biomization method, and many of the more systematic errors were eliminated. The pollen-derived biomes match observed biomes in

extent, in internal cohesiveness, and have improved kappa scores.

Tundra is the least affected by these modifications, and has the lowest individual kappa score ($\kappa_T = 0.58$). Of the 26 sites in the modern tundra, only 13, all of which are east of Hudson Bay, are correctly classified (Table 3c). The other thirteen sites, however, are mostly within 50 km of observed boreal forest, and are classified as taiga.

Ninety-one percent of the sites in the observed boreal forest are classified as taiga, and the taiga biome has the best individual kappa score ($\kappa_{Ta} = 0.89$). Five sites are classified as tundra (Table 3c), most of which occur within 15 km of observed tundra. The remaining 12 mismatches are cool mixed forest sites, which mostly fall within 50 km of observed cool mixed forest. Not included in the calculation of κ_{Ta} and hit frequency are 19 sites classified as cool conifer forest, occurring along the southern fringe of observed boreal forest where cool-temperate taxa such as *Ulmus*, *Fagus*, *Acer*, and *Tsuga* are present. If these sites are included, κ_{Ta} drops to 0.83, still a very good score.

Sites in the observed mixed forest are well characterized by the biomization method as representing cool mixed forest ($\kappa_{CM} = 0.79$); of the 243 sites occurring in the mixed forest, 94% are classified as cool mixed forest or cool conifer (Table 7). Cool conifer forest, not considered as a category by Overpeck *et al.* (1992), is most similar to cool mixed forest (Table 2), and is distinguished from cool mixed forest by an absence of temperate summergreens (e.g. *Quercus*, *Carya*, and *Fraxinus*). The cool conifer sites occur in a swath from southern Manitoba to Nova Scotia and southwestern Newfoundland. Olson *et al.* (1984) also mapped cool conifer forest throughout most of this area; however, between the Great Lakes and the St. Lawrence River the cool conifer forest zone predicted by the biomization method extends approximately 150 km south of the zone mapped by Olson *et al.* (1984).

Deciduous forest has a good to very good kappa score ($\kappa_{TD} = 0.71$), with 77% of the 197 sites classified as temperate deciduous forest. Of the mismatches, 38 of 46 are classified as cool mixed forest (Table 3c). Most of these cool mixed sites occur in three areas: (1) within 50 km of observed cool mixed forest, (2) in the Appalachian Mountains where higher resolution maps show cool mixed forest (Fowells, 1975; Olson *et al.*, 1984; Webb, 1987), and (3) on Cape Cod where *Pinus* is abundant due to sandy soils (Winkler, 1985).

Seventy-seven percent of sites in the southeastern forest are classified as warm mixed forest by the biomization method ($\kappa_{WM} = 0.81$). Of the remaining sites, 18 of 21 are classified as temperate deciduous forest (Table 3c). Of the temperate deciduous sites, three occur within 50 km of the observed deciduous forest/southeastern forest border, and 11 occur within a corridor along the Mississippi River observed to be river bottom forest and non-forest vegetation (Fowells, 1975). River bottom forest is predominantly Cypress-Tupelo-Sweetgum (*Taxodium-Nyssa-Liquidambar*), and

TABLE 7. Weighted averages of the percent agreement between the biomization method and either the modern analog technique (MAT) or BIOME1

	MODERN				6 KA			
	Modern Potential Biomes vs.				BIOME1 vs.	MAT vs.		BIOME1 vs.
	1st iteration (Fig. 1)	2nd iteration (Fig. 3)	3rd iteration (Fig. 4)	'Optimal'	3rd iteration (Fig. 6)	3rd iteration (Fig. 5)	'Optimal'	3rd iteration (Fig. 7)
Overall score	0.73	0.80	0.85	0.92	0.33	0.85	0.92	0.26
Standard deviation	0.03	0.03	0.03	0.03	0.03	0.05	0.04	0.06
Tundra	0.35	0.50	0.50	0.50	0.35	1.00	1.00	1.00
Taiga	0.63	0.60	0.91	0.96	0.97	0.87	0.93	0.63
Cool Conifer Forest	—	—	—	—	0.20	—	—	0.15
Cool Mixed Forest	0.77	0.97	0.94	0.97	0.57	0.91	0.95	0.69
Temp. Deciduous Forest	0.77	0.68	0.77	0.86	0.68	0.64	0.79	0.35
Warm Mixed Forest	0.73	0.71	0.77	0.92	0.80	0.82	0.90	0.75
Steppe	0.39	0.61	0.67	0.69	0.92	0.83	0.89	1.00

the low levels of *Pinus* (usually <10%) at these sites causes the biomization method to choose temperate deciduous forest over warm mixed forest.

Steppe has a very good kappa score ($k_s = 0.72$) and of the 18 sites in steppe 12 are correctly assigned by the biomization method. The six remaining sites, occurring from eastern South Dakota to northern Missouri, are mostly less than 100 km from observed temperate deciduous forest. These sites are presumably influenced by arboreal pollen transported from the east or arboreal pollen from local tree-rich drainages. These sites typically contain a minimum of 30–40% forb pollen, underscoring the point that the biomization method in general is more sensitive to the presence/absence of taxa rather than to large abundances.

The resulting biome map (Fig. 3) is highly similar to the observed and potential biome distributions, with only the prairie-forest boundary remaining significantly offset. The corresponding kappa score (Table 6), however, is more conservative, as it captures discrepancies not only caused by errors in the biomization method but from other causes as well. First, individual pollen samples may capture the intermingling nature of most biome boundaries if the pollen site is within a localized patch of an adjacent biome, but the modern observed biome transitions are mapped as abrupt contacts. Furthermore, even when vegetative gradients are fairly abrupt, the dispersive nature of pollen acts to smooth the pollen gradients (Jackson *et al.*, 1997). Second, the limited number of biomes and the relatively coarse resolution of the Overpeck maps masks smaller-scale differences in the vegetation. The Appalachian Mountain sites classified as cool mixed forest are one example; another is the river bottom forest occurring along the southern Mississippi River

(Fowells, 1975), which is included within southeastern forest by Overpeck *et al.* (1992). Counting sites identified as temperate deciduous forest in this region as evidence against the accuracy of the biomization method does not take into account compositional differences between river bottom forest and warm mixed forest. River bottom forest is predominated by *Taxodium*, *Nyssa*, and *Liquidambar* and has very low *Pinus* abundances, while warm mixed forest is composed mostly of *Pinus* and *Quercus*.

To provide a kappa score that did not penalize the biomization method for other sources of error, we deleted the 16 sites within the river bottom forest (as defined by Fowells, 1975) and the 204 sites within 50 km of a border and recalculated the kappa statistic with the remaining 572 sites. We chose 50 km because the precision of potential biome maps, as estimated by comparing those in Webb (1987) and Overpeck *et al.* (1992), is approximately 100 km. As a result of eliminating these 220 sites, the kappa score improves from 0.79 to 0.89, suggesting that the agreement between the biomes inferred from the biomization method and those observed may be better than our conservative estimate (Table 6).

In summary, after modification, the biomization method accurately classifies ENA pollen data into biomes. Along most biome boundaries there is some blurring, as might be expected from observed transitions between biomes, but only along the steppe/forest boundary is there a systematic bias. This overall success can be further evaluated by comparing the reconstruction of 6 ka vegetation by the biomization method to that by the MAT (Overpeck *et al.*, 1992) to see whether the modifications link the biomization method too tightly to present-day patterns.

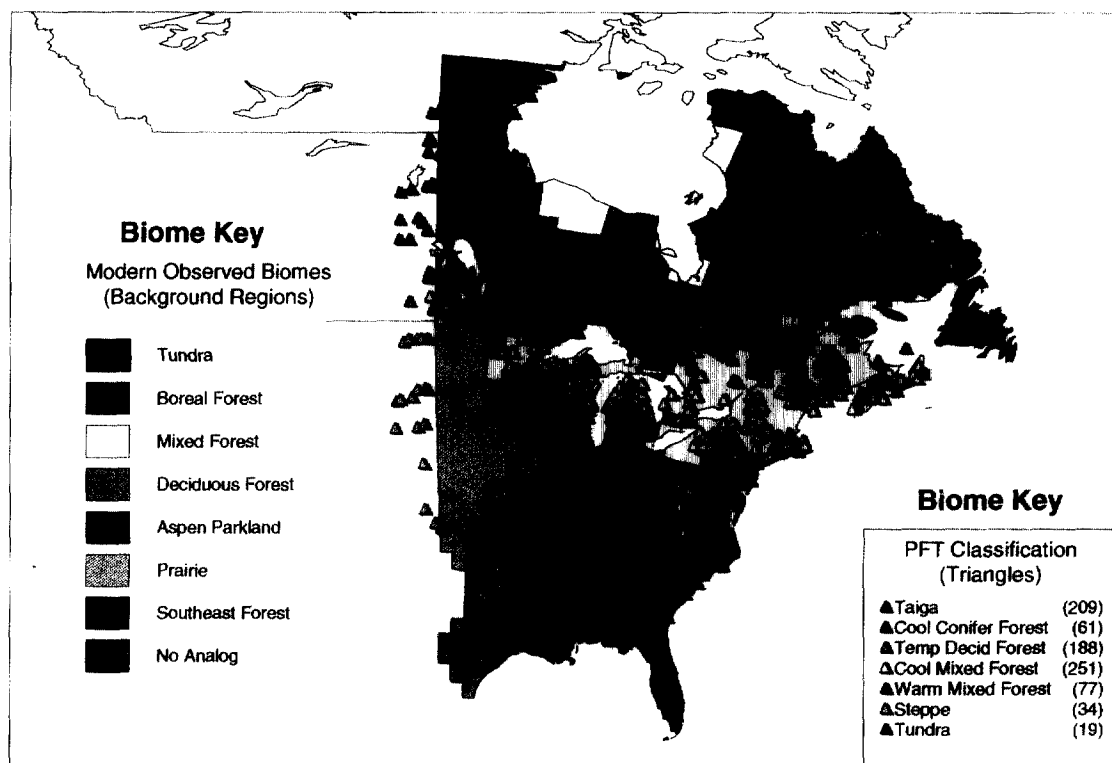


FIG. 3. The third iteration of the biomization method overlain upon modern observed and potential biomes. See the caption for Fig. 1 for a fuller description.

COMPARING BIOMIZATION METHOD AND MAT RESULTS AT 6 ka

The biome maps produced by the biomization method for 6 ka can be evaluated by comparison to biome maps inferred by the MAT to the biome map produced for Quebec by Richard (1995), and to the movement of plant taxa as recorded by shifts in pollen and macrofossil frequencies. The biomization method and MAT vegetation maps for 6 ka (Fig. 4) agree well qualitatively (Fig. 4) and quantitatively (Tables 6 and 7). Removing border sites as described above (the presence of river bottom forest was not considered for 6 ka) raises the kappa score from 0.75 (very good) to 0.87 (excellent). The two maps show no serious discrepancies or offsets in biome distributions, including the steppe/forest boundary. That the biome shifts inferred by both methods are in fact matched by movements of pollen taxa is supported by isopoll maps (Webb, 1987; Webb *et al.*, 1993), and the pollen record is further supported by similar spatial patterns in the macrofossil record (Jackson *et al.*, 1997). Both biome maps show a northward extension of the cool mixed forest and an eastward movement of prairie in the Midwest — two long-recognized patterns in pollen maps (Webb *et al.*, 1983a, b; Wright, 1968).

All seven sites in the region classified as tundra by the MAT are classified by the biomization method as tundra ($\kappa_T = 0.82$), and match other pollen-based reconstructions of the 6 ka tundra-taiga boundary

(Richard, 1995). Areas classified as boreal forest by the MAT show a close match to sites classified as taiga by the biomization method, with a kappa score of 0.85. The exceptions are four sites classified as tundra and five sites classified as cool conifer forest (Table 3b); if these cool conifer sites are included in the calculations, κ_{Ta} drops to 0.79.

The region identified by the MAT as mixed forest is well characterized ($\kappa_{CM} = 0.75$) by the biomization method, with 92% classified as cool mixed or cool conifer forest (Table 3b). The extent of cool conifer forest at 6 ka is similar to its inferred extent today. The few areas identified as containing no-analog vegetation by the MAT are classified as cool mixed forest by the biomization method. The appearance of three warm mixed forest sites in the northern mixed forest is caused by a combination of high abundances of *Pinus* with an absence of northern types such as *Picea*, *Abies*, and *Tsuga* that would indicate cool mixed forest over warm mixed forest. Clearly a difficulty here is distinguishing between northern and southern pines — a problem that may be overcome by referring to the macrofossil record (Jackson *et al.*, 1997) or by classifying *Pinus* pollen on the basis of the accompanying pollen taxa (Webb *et al.*, 1993).

The region classified by the MAT as deciduous forest is matched well by the temperate deciduous forest sites identified by the biomization method. The kappa score (0.64) is relatively low, however, because of a number of cool mixed forest sites in southern New

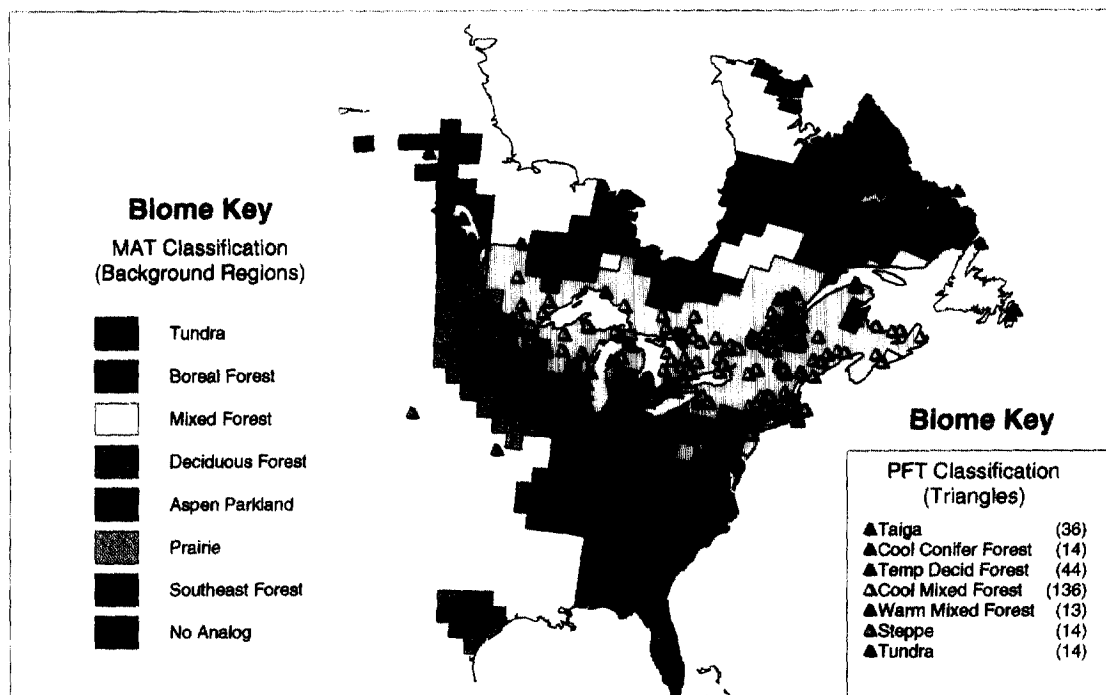


FIG. 4. The third iteration of the biomization method now applied to pollen data from 6 ka, overlain upon a biome map obtained by applying the MAT to fossil pollen data (Overpeck *et al.*, 1992). The color scheme is the same as in Fig. 1.

England. Of the 19 cool mixed forest sites identified in the MAT-inferred deciduous forest, five occur along the Appalachians and ten are within 40 km of MAT mixed forest. Nine of the 11 sites occurring in the MAT-inferred southeastern forest are classified as warm mixed forest ($\kappa_{WM} = 0.74$).

The movement of biome boundaries inferred by the biomization method and MAT can be explained as the sum of individual plant migrations. The major features of the 6 ka biome map relative to today are (1) a more southerly extent of tundra in Quebec (Richard, 1995), (2) a reduced taiga and more northerly cool mixed forest, (3) a stable temperate deciduous-cool mixed forest boundary, (4) an eastward shift of the steppe-forest boundary, and (5) a reduced warm mixed forest. Each of these features can be linked to the shifts of one or more major plant types as indicated in both the pollen and macrofossil records (Webb *et al.*, 1993, Jackson *et al.*, 1997). The reduced extent of taiga and the corresponding encroachments by tundra and cool mixed forest is signaled by a contraction of the distribution of *Picea* pollen and a northward shift of *Pinus*, and *Tsuga* pollen. An eastward shift in the Cyperaceae and prairie forb isopolls is echoed by an eastward but lesser movement of the prairie-forest boundary. The reduction in the warm mixed forest at 6 ka is explained by the more limited range of the southern pines.

Overall the biome maps produced by the biomization method show a close similarity to maps not only of modern observed and potential vegetation but also of past vegetation as reconstructed by the MAT. The

agreement, although not excellent, is consistently very good, suggesting that the modified biomization method and the MAT are perceiving consistent and valid patterns from the pollen data. Furthermore, the inferred movements of biomes from 6 ka to today are matched by the shifts in isopolls of individual taxa (Webb *et al.*, 1993; Webb *et al.*, 1983a, b; Wright, 1968), and the pollen record is itself supported by the available macrofossil data (Jackson *et al.*, 1997). In the next section, we compare the vegetation patterns inferred from the biomization method to vegetation simulated by BIOME1 from CCM1 (Kutzbach *et al.*, 1998), and test whether BIOME1 will match the biomization method as well as the latter matched modern vegetation and MAT-inferred vegetation.

EVALUTATING BIOME1 RESULTS WITH BIOMES FROM POLLEN DATA

Biome Estimation for Today

For today's vegetation, BIOME1 mapped most biomes in locations generally similar to those of the biomization method, but the corresponding kappa score ($\kappa = 0.53$) indicates only a fair agreement (Fig. 5; Tables 6 and 7). BIOME1 correctly captures the north to south sequence of biomes and roughly approximates their sizes, but the locations of biome boundaries are offset significantly in several places. In general, wherever BIOME1 and the biomization method agree, both match observed vegetation, but where they disagree

BIOME1 also notably differs from observed biomes. The closest agreement is seen in the warm mixed forest ($\kappa_{WM} = 0.77$), which has a very similar distribution under the two classification schemes. Taiga also is in good visual agreement, although the biomization method infers taiga to extend farther north than does BIOME1.

Differences, however, predominate, of which seven stand out. First and foremost is the offset of the cool mixed/temperate deciduous forest boundary, which BIOME1 places too far south. The boundary is increasingly offset to the west, up to as much as 600 km. The placement of this boundary greatly affects kappa scores because of the high density of pollen sample sites in this area. Second, BIOME1 simulates no cool mixed forest in northern Minnesota west of Lake Superior in contrast to the modern vegetation and pollen evidence for this biome there (Fig. 5). Third, BIOME1 simulates cool conifer forest mostly north of where the biome is placed by the biomization method. The two regions of cool conifer agree near Lake Winnipeg, but from there eastwards they diverge, to the point of no overlap in the Quebec–New Brunswick–Maine region. The mapping of observed cool conifer forest by Olson *et al.* (1984) agrees partially with both. Olson *et al.* (1984) map cool conifer forest in Maine and New Brunswick (agreeing with the biomization method) but farther west, they map cool conifer forest in southern Ontario

and Quebec (agreeing with BIOME1). Fourth, tundra east of Hudson Bay is placed too far south by BIOME1 relative to the biomization method and is farther south than shown by the map of modern vegetation. Fifth, although steppe has a high kappa score ($\kappa_S = 0.85$), BIOME1 places the steppe/forest boundary somewhat west of the boundary set by the biomization method, which is in turn west of the observed boundary (Figs 3 and 5). Sixth, the biomization method did not predict any regions of cold deciduous or cold mixed forest. This is because we removed these biomes from consideration, but even when we performed experiments (not shown here) allowing cold deciduous and cold mixed forest to be considered the two biomes were not represented to any significant degree. BIOME1 predicts a small region of cold deciduous and cold mixed forest south and west of Lake Winnipeg, but the biomization method infers instead cool mixed forest and steppe due to the presence of boreal evergreen conifer, prairie forb, and occasional temperate summergreen pollen in this region. Seventh, tropical forest, a biome not considered in our use of the biomization method, is simulated by BIOME1 for southern Florida. Fowells (1975) reports southeastern pine forest approximately as far south as the southernmost pollen site (where the biomization method infers a warm mixed forest), south of which they report non-forest vegetation.

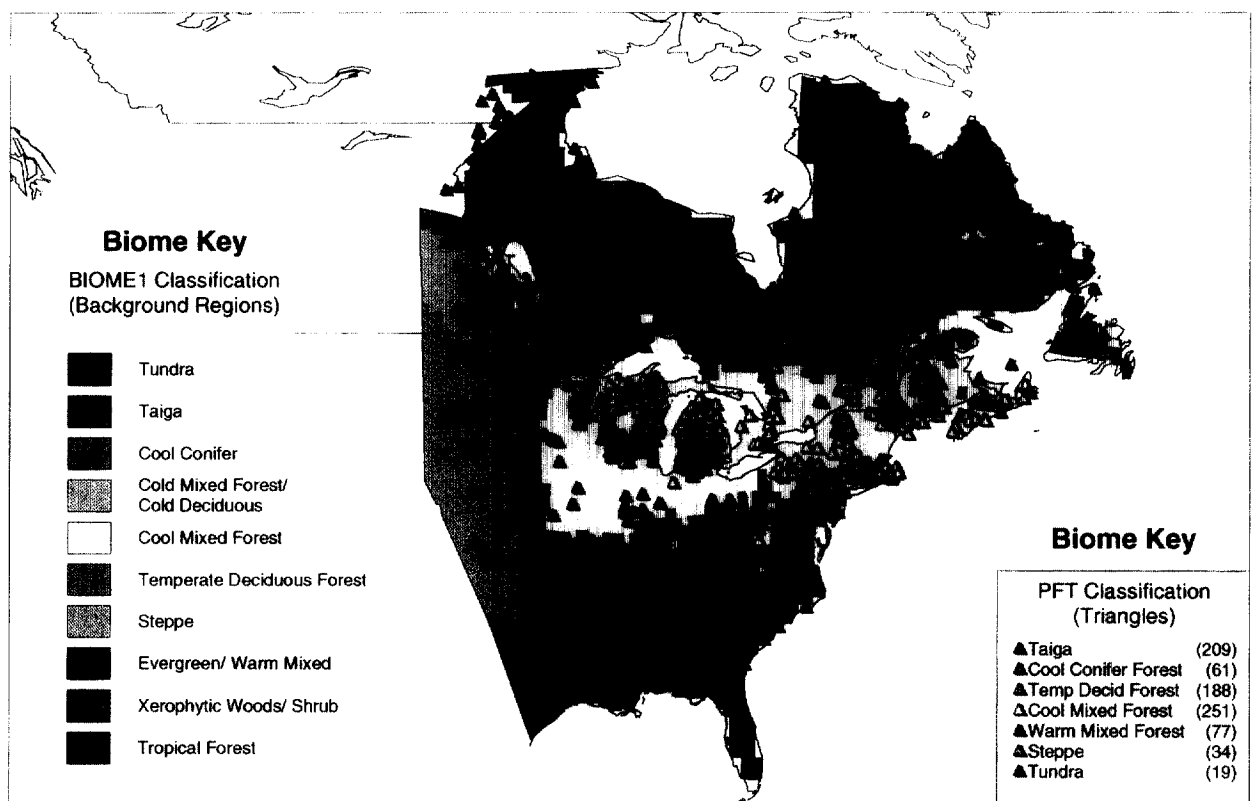


FIG. 5. The third iteration of the biomization method, applied to modern pollen data, overlain upon biome regions simulated by BIOME1 from the reconstruction of modern climate by CCM1. A slightly expanded color key is used for the BIOME1 biomes to accommodate its additional biomes.

Biome Estimation for 6 ka

Most of the similarities and discrepancies identified between the results of the biomization method and BIOME1 for 0 ka persist at 6 ka (Fig. 6; Tables 6 and 7). BIOME1 again captures the north-south sequence of biomes, but greater offsets between the data and the model are reflected in increasingly divergent biome locations and boundaries, and the kappa score drops to 0.39, indicating a poor level of agreement overall. As for today, warm mixed forest shows the most similarity between the two reconstructions, and the distribution of taiga also matches fairly well.

The boundary between the cool mixed and temperate deciduous forests lies perpendicular to the boundary inferred by the biomization method. The northwestern region of cool mixed forest inferred by both the biomization method and the MAT is not simulated by BIOME1, which places cool mixed forest much farther southwest. BIOME1 consequently simulates temperate deciduous forest to have a much smaller area than inferred by the biomization method.

Simulated cool conifer forest is offset from inferred cool conifer forest as much at 6 ka as at 0 ka. Both methods agree south of Hudson Bay, but farther west BIOME1 extends cool conifer forest southwest of Lake Superior, an area identified as cool mixed and temperate deciduous forest by the biomization method and the MAT.

Simulated tundra does not extend as far south as tundra inferred by the biomization method and MAT, which in turn is supported by a similar placement of the tundra-taiga boundary by Richard (1995). The BIOME1 steppe/forest boundary is significantly farther west than is inferred by the biomization method and MAT. BIOME1 predicts cold deciduous forest along the margins of steppe, but the few sites in this region are classified by the biomization method as steppe or temperate deciduous forest. As at modern, the biomization method predicts warm mixed forest where BIOME1 predicts tropical forest.

Discrepancies can arise from flaws in the biomization method, BIOME1, or CCM1. Because the biomes reconstructed by the biomization method are closer to the MAT and other interpretations of the paleovegetation data than are the simulated biomes, discrepancies are likely due to biases in BIOME1 or CCM1 (see below).

DISCUSSION

The goal of this study was to adapt the biomization method for use with eastern North American pollen data, and to use the pollen-derived biome maps to evaluate the biome simulations of BIOME1 (and, indirectly, the climate simulations of CCM1). We found that the biomization method successfully reproduced modern and 6 ka biome patterns but only after certain

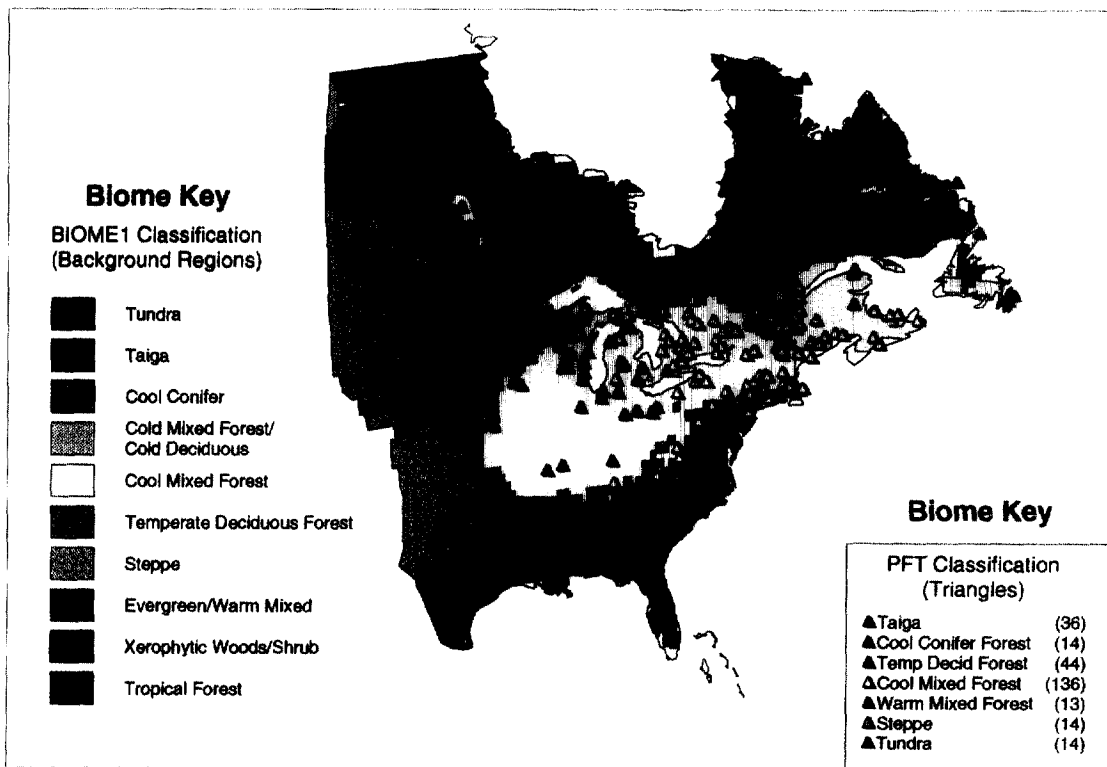


FIG. 6. The third iteration of the biomization method, applied to fossil pollen data from 6 ka, overlain upon biome regions simulated by BIOME1 from the 6 ka climate reconstructions of CCM1. The color scheme is the same as for Fig. 5.

modifications were made to what Prentice *et al.* (1996) did (e.g. selectively raising cutoff values and adjusting PFT–biome assignments). In this section, we review the results of the biomization method and the modifications made, and discuss whether the need for modifications to the biomization method is an inherent consequence of the nature of pollen data, or instead is driven by the relatively limited number of pollen taxa used in this study. We also weigh the relative advantages of the biomization method and the MAT for reconstructing biomes from pollen data and evaluating biome simulations, and discuss possible causes for the discrepancies between the biome maps of BIOME1 and the biomization method.

Maps of ENA biomes reconstructed using the modified biomization method agree well with observed modern vegetation ($\kappa = 0.79$) and maps of vegetation at 6 ka ($\kappa = 0.75$) as inferred by the MAT, and are consistent with the spatial patterns of individual pollen abundances and macrofossil distributions (Figs 3 and 4; Tables 7 and 8; Webb *et al.*, 1993; Jackson *et al.*, 1997). This success gives confidence that biome maps reconstructed by the biomization method accurately represent the actual vegetation and may be used in turn to test other biome reconstructions. The one significant exception is the steppe/forest boundary, which the biomization method placed approximately 100 km too far to the west relative to modern potential vegetation, and which both the biomization method and the MAT place too far west when compared with isopoll data. The bias in the steppe/forest boundary may be caused by the greater mobility of arboreal pollen relative to herb pollen; Davis and Webb (1975) observe that tree pollen can occur in significant amounts (>20%) in treeless areas. Alternatively, the eastern steppe may have been closer to a savanna, with scattered trees providing local sources of arboreal pollen. Because the height of trees make them more efficient pollen dispersers, even a minority population could outweigh the pollen production of the herbs. To address these possibilities, we plan to add Poaceae to our taxa list in the hope of improving the representation of steppe, and are experimenting with the addition of a parkland or savanna biome.

A major rationale for developing the biomization method was to have a means of identifying biomes flexible enough to handle the diverse array of plant assemblages found globally, which is achieved by coupling a global list of PFT–biome assignments with regional assignments of plant taxa to PFTs (Prentice *et al.*, 1996). Any modifications made to the biomization method to fit regional patterns in the pollen data beyond altering taxon–PFT assignments potentially weakens the global nature of the method. We nevertheless found it necessary to modify the biomization method in order to achieve a reasonable fit to modern ENA biomes; these modifications fall into four categories: (1) defining a new tie break procedure, (2) raising the general threshold value from 0.5 to 1%, (3) changing PFT–biome assignments, and (4) raising the threshold

value for selected pollen taxa. The first and second modifications do not affect the generality of the biomization method, but their impact should be evaluated for other regions. Adjusting PFT–biome assignments was necessary given that the assignments set by Prentice *et al.* (1996) for European PFTs do not completely match the observed distributions of PFTs in ENA biomes. In particular, in ENA, conifers are only present in limited quantities in the temperate deciduous forest and Cyperaceae is found in the steppe. Our PFT–biome assignments are not intended to replace the assignments of Prentice *et al.* (1996), but are proposed in order to point out the differences between the PFT–biome relationships in ENA and Europe. These differences must be reconciled before a PFT–biome list can be considered globally applicable.

Raising thresholds selectively is problematic. On the one hand, altering thresholds provides a simple means of accounting for over- and under-representation of plant taxa in the pollen record. Selectively raising thresholds in order to improve the results of the biomization method, on the other hand, reduces the generality of the biomization method. The exact level of the threshold needed will vary from region to region, and will even be affected by the choice of the pollen sum. Our need to modify thresholds arises in part because of our reliance on a limited number of taxa; examination of Table 5 reveals that biomes are frequently differentiated only by one or two taxa, whose presence or absence becomes critical in choosing between closely related biomes. Increasing the number of taxa should reduce the dependence on any given taxon, and should lessen the need for selectively raised thresholds. Even with more taxa added, however, it is likely that some biomes will remain subsets of others (e.g. taiga, cool conifer forest, cool mixed forest), in which case a need for selectively raised thresholds remains, because the biomization method is highly sensitive to the presence of the differing taxa. For example, a pollen assemblage composed of *Picea*, *Abies*, *Pinus*, *Betula*, *Alnus*, and/or *Populus*, and less than 1% *Quercus*, will classify as taiga (see Table 5), but a pollen assemblage identical to the first in all ways but containing 1.1% *Quercus* will classify as cool mixed forest. Yet a low but >1% pollen abundance may not necessarily indicate even a minor presence of a taxon on the landscape because the pollen of some taxa, such as *Pinus* and *Quercus*, can be transported long distances (Webb *et al.*, 1981). We believe that in the future using more taxa will reduce but not eliminate the need for taxon-specific thresholds.

The question may be raised as to why we choose to use the biomization method when the MAT is already an established means of inferring biomes to pollen data. Both are viable means of reconstructing biomes from pollen data, each having its strengths and weaknesses, and can be highly complementary to one another. The MAT is more sensitive to pollen taxa with high abundances than the biomization method, which

is affected more by the presence or absence of a taxon. The advantages of the biomization method over the MAT are twofold. First, the biomization method does not rely on large sets of modern samples for calibration and is therefore useful for areas for which few sites are available. Because the biomization method does not require an extensive modern data set, although one is useful for validation, biomes may be defined regardless of whether they are present in the modern landscape. Thus the biomization method can be applied to times with vegetation assemblages that have no modern counterparts. Third, the biomization method is structurally similar to BIOME1 in that both use plant functional types as their basic vegetational units and define biomes as assemblages of PFTs, which makes the pollen-derived biomes of the biomization method particularly useful for assessing the biome simulations of BIOME1. The major disadvantage of the biomization method relative to the MAT is that it must assign every pollen sample to a biome, potentially lumping dissimilar pollen samples into the same biome. The MAT and biomization method thus complement one another for pollen samples that have no modern analog. The MAT cannot classify such samples, but identifies them as having no modern analog. The biomization method would not have identified the pollen samples as fundamentally different, but, once identified, new biome definitions may be added into the biomization method to accommodate these no analog biomes.

Data/Model Comparisons

When the biomization method was used to evaluate the vegetation simulations of BIOME1, the overall agreement between the model results and the data was only fair to poor, and several specific discrepancies existed. The major shortfalls in the BIOME1 results are its location of the temperate deciduous/cool mixed forest boundary and the steppe/forest boundary, which were consistently south or west of the data both for today and 6 ka. These discrepancies may have been due to inaccuracies in the climate simulation (CCM1), the simulation of biomes from climate data (BIOME1) or the classification of pollen data (biomization method), but the accuracy of the biomization method in comparison with modern observations and the MAT gives confidence that the biomization method is not the major source of error. Webb *et al.* (1997) applied pollen response surfaces to infer climate variables from a similar pollen data set for 0 and 6 ka and provided a direct assessment of the CCM1 climate simulations for ENA. The regional discrepancies in climate and pollen terms between simulated and inferred patterns at 6 ka are small enough that the simulated biomes should match the pollen-derived biomes as well as the PFT-derived biomes match the MAT biomes at 6 ka. That they do not indicates that biases in BIOME1 cause the general problems that BIOME1 showed in matching the results of the biomization

method. BIOME1 needs to place the cool mixed/temperate deciduous forest boundary farther north in the Midwest, to simulate cool mixed forest in northern Minnesota, and to simulate the steppe/forest border farther east. These problems are being addressed by modeling improvements in BIOME3 (Haxeltine and Prentice, 1996).

We are continuing to work with the biomization method. Current and future investigations include (1) incorporating more taxa from the pollen record and splitting up groups like prairie forbs to decrease the need to rely on taxon-specific thresholds, (2) extending the method to other time slices, and defining new biomes to account for the no analog vegetation of the late Pleistocene, (3) reapplying the biomization method back to the European data base to assess the impact of our modifications, and (4) estimating past and present carbon storage for ENA from our vegetation maps.

ACKNOWLEDGEMENTS

A grant from the NSF Earth System History program to TEMPO (Testing Earth System Models with Paleoenvironmental Observations) and a DOE grant from the Office of Energy Research supported this research. We thank I.C. Prentice for information about the biomization method and also for his comments on the manuscript. Helpful reviews were also provided by S. Jackson and P. Richard. We are grateful to P. Newby for her help and to K. Anderson, P. Leduc and A. Rosen for technical support. We thank J. Kutzbach and S. Harrison for providing the BIOME1 maps and D. McClure for advice on statistics.

REFERENCES

- Adams, J.M., Faure, H., Faure-Denard, L., McGlade, J.M. and Woodward, F.I. (1990) Increases in terrestrial carbon storage from the Last Glacial Maximum to the present. *Nature*, **348**, 711–714.
- Arai, K. (1993) A classification method with a spatial-spectral variability. *International Journal of Remote Sensing*, **14**(4), 699–709.
- Bard, E., Hamelin, B. and Fairbanks, R.G. (1990) U–Th ages obtained by mass spectrometry in corals from Barbados: sea level during the past 130,000 years. *Nature*, **346**, 456–458.
- Barnola, J.M., Raynaud, D., Korotkevich, Y.S. and Lorius, C. (1987) Vostok ice core provides 160,000-year record of atmospheric CO₂. *Nature*, **329**, 408–414.
- Bartlein, P.J., Prentice, I.C. and Webb III, T. (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, **13**, 35–57.
- Bartlein, P.J., Anderson, K.H., Anderson, P.M., Edwards, M.E., Mock, C.J., Thompson, R.S., Webb, R.S., Webb, T. III and Whitlock, C. (1998) Paleoclimate simulations for North America over the past 21,000 years: features of the simulated climate and comparisons with paleoenvironmental data. *Quaternary Science Reviews* **17**, 549–585.
- Bernabo, J.C. and Webb, T. III (1977) Changing patterns in the Holocene pollen record of northeastern North America: A mapped summary. *Quaternary Research*, **8**, 64–96.

- Bradshaw, R.H.W. and Webb, T. III (1985) Relationships between contemporary pollen and vegetation data from Wisconsin and Michigan, USA. *Ecology*, **66**, 721–737.
- Clements, F.E. and Shelford, V.E. (1939) *Bio-ecology*. Wiley, New York.
- COHMAP Members (1988) Climatic changes of the last 18,000 years: observations and model simulations. *Science*, **24**, 1043–1052.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**(1), 37–46.
- Davis, R.B. and Webb, T. III (1975) The contemporary distribution of pollen in eastern North America: A comparison with the vegetation. *Quaternary Research*, **5**, 395–434.
- Delcourt, P.A. and Delcourt, H.R. (1981) Vegetation maps for eastern North America: 40,000 yr B.P. to the present. In: R.C. Romans (Eds), *Geobotany II*. Plenum Press, New York.
- Fægri, K., Kaland, P.E. and Krzywinski, K. (1989) *Textbook of Pollen Analysis*. Wiley, New York, NY.
- Fowells, H.A. (1975) *Silvics of the United States*. U.S. Department of Agriculture, Washington, DC.
- Frenzel, B. (1973) *Climatic Fluctuations of the Ice Age*. Press of Case Western Reserve University, Cleveland, OH.
- Frenzel, B., Pecsli, M. and Velichko, A.A. (1992) *Atlas of Paleoclimates and Paleoenvironments of the Northern Hemisphere*. Gustav Fischer Verlag, Stuttgart.
- Guetter, P.J. and Kutzbach, J.E. (1990) A modified Koppen classification applied to model simulations of glacial and interglacial climates. *Climate Change*, **16**, 193–215.
- Haxeltine, A. and Prentice, I.C. (1996) BIOME3: an equilibrium biosphere model based on ecophysiological constraints, resource availability and competition among plant functional types. *Global Biogeochemical Cycles*, **10**(4), 693–709.
- Jackson, S.T., Overpeck, J.T., Webb, T. III, Keatch, S.E. and Anderson, K.H. (1997) Mapped plant-macrofossil and pollen records of late Quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, **15**, 1–71.
- Jacobson, G.L. and Bradshaw, R.H.W. (1981) The selection of sites for paleoenvironmental studies. *Quaternary Research*, **16**, 80–96.
- Kuchler, A.W. (1964) *Potential Natural Vegetation of the Conterminous United States*. Princeton Polychrome Press, New York.
- Kutzbach, J., Gallimore, R., Harrison, R., Behling, P., Selin, R. and Laarif, F. (1998) Climate and biome simulations for the past 21,000 years. *Quaternary Science Reviews* **17**, 473–506.
- Lorius, C., Raynaud, D., Petit, J.-R., Jouzel, J. and Merlivat, L. (1984) Late-glacial maximum-Holocene atmospheric and ice-thickness changes from Antarctic ice-core studies. *Annals of Glaciology*, **5**, 88–94.
- Monserud, R.A. (1990). *Methods for Comparing Global Vegetation Maps*. IIASA, Laxenburg.
- Monserud, R.A. and Leemans, R. (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, **62**, 275–293.
- Nezry, E., Mougín, E., Lopes, A., Gastellu-Etchegorry, J.P. and Laumonier, Y. (1993) Tropical vegetation mapping with combined visible and SAR spaceborne data. *International Journal of Remote Sensing*, **14**, 2165–2184.
- Oeschger, H., Beer, J., Siegenthaler, U., Stauffer, B., Dansgaard, W. and Langway, C.C. (1983) Late-glacial climate history from ice cores. In: Ghazi, A. (ed.), *Paleoclimatic Models and Research*, pp. 95–107. Reidel, Dordrecht.
- Olson, J.S., Watts, J.A. and Allison, L.J. (1984). *Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: a Database*. Oak Ridge National Laboratory, Oak Ridge.
- Overpeck, J.T. (1985) A pollen study of a late Quaternary peat bog, south-central Adirondack Mountains, New York. *Geological Society of America Bulletin*, **96**, 145–154.
- Overpeck, J.T., Webb, T. III and Prentice, I.C. (1985) Quantitative interpretation of fossil pollen spectra: dissimilarity coefficients and the method of modern analogs. *Quaternary Research*, **23**, 87–108.
- Overpeck, J.T., Webb, R.S. and Webb, T. III (1992) Mapping eastern North American vegetation change of the past 18 ka: No-analogs and the future. *Geology*, **20**, 1071–1074.
- Peltier, W.R. (1994) Ice age paleotopography. *Science*, **265**, 195–201.
- Prentice, I.C. (1988) Records of vegetation in time and space: the principle of pollen analysis. In: Huntley, B. and Webb, T. III (eds), *Vegetation History*, pp. 17–42. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. and Solomon, R.A. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117–134.
- Prentice, I.C., Guiot, J., Huntley, B., Jolly, D. and Cheddadi, R. (1996) Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Climate Dynamics*, **12**, 185–194.
- Prentice, I.C. and Webb, T. III. (1998) BIOME 6000: reconstructing global mid-Holocene vegetation patterns from palaeoecological records. *Journal of Biogeography* (in press).
- Ranson, K.J., Saatchi, S. and Sun, G. (1995) Boreal forest ecosystem characterization with SIR-C/XSAR. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 867–876.
- Ranson, K.J. and Sun, G. (1994) Northern forest classification using temporal multifrequency and multipolarimetric SAR images. *Remote Sensing of the Environment*, **47**, 142–153.
- Richard, P.J.H. (1995) Le couvert végétal du Québec-Labrador il y a 6000 ans BP: essai. *Géographie physique et Quaternaire*, **49**, 117–140.
- Rowe, J.S. (1972) *Forest Regions of Canada*. Canadian Forestry Service.
- Stuiver, M. and Reimer, P.J. (1993) Extended ¹⁴C data base and revised CALIB 3.0 ¹⁴C age calibration program. *Radiocarbon*, **35**, 215–230.
- VEMAP Members (1995) Vegetation/ecosystem modeling and analysis project (VEMAP): comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and CO₂ doubling. *Global Biogeochemical Cycles*, **9**, 407–437.
- Webb, T. III (1987) The appearance and disappearance of major vegetational assemblages: Long-term vegetational dynamics in eastern North America. *Vegetatio*, **69**, 177–187.
- Webb, T. III (1988) Eastern North America. In: Huntley, B. and Webb T. III (ed.), *Vegetation History*, pp. 385–414. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Webb, T. III (1993) Constructing the past from late-Quaternary pollen data: temporal resolution and a zoom lens space-time perspective. In: Kidwell, S.M. and Behrensmeier, A.K. (eds), *Taphonomic Approaches to Time Resolution in Fossil Assemblages*, pp. 79–101. Paleontological Society Short Courses in Paleontology No. 6. University of Tennessee, Knoxville, TN.
- Webb, T. III, Anderson, K.H., Bartlein, P.J. and Webb, R.S. (1998) Late Quaternary climate change in eastern North America: a comparison of pollen-derived estimates with climate model results. *Quaternary Science Reviews* **17**, 587–606.
- Webb, T. III, Howe, S.E., Bradshaw, R.H.W. and Heide, K.M. (1981) Estimating plant abundances from pollen percentages: the use of regression analysis. *Review of Palaeobotany and Palynology*, **34**, 269–300.
- Webb, T. III, Richard, P.J.H. and Mott, R.J. (1983a) A mapped history of Holocene vegetation in southern Quebec. *Syllogeus*, **49**, 273–336.
- Webb, T. III, Cushing, E.J. and Wright, H.E., Jr. (1983b) Holocene changes in the vegetation of the midwest. In: Wright, H.E., Jr. (eds), *Late-Quaternary Environments of the United States*, pp. 142–165. University of Minnesota Press, Minneapolis, MN.

Webb, T. III, Bartlein, P.J., Harrison, S.P. and Anderson, K.H. (1993) Vegetation, lake levels, and climate in eastern North America for the past 18,000 years. *In*: Wright, H.E., Jr, Kutzbach, J.E., Webb, T. III, Ruddiman, W.F., Street-Perrott, F.A. and Bartlein, P.J. (eds), *Global Climates since the Last Glacial Maximum*, pp. 415–467. University of Minnesota Press, Minneapolis, MN.

Winkler, M.G. (1985) A 12,000-year history of vegetation and climate for Cape Cod, Massachusetts. *Quaternary Research*, **23**, 301–312.

Wright Jr., H.E. (1968) The roles of pine and spruce in the forest history of Minnesota and adjacent areas. *Ecology*, **49**, 937–955.