Numerical Analysis of Modern and Fossil Pollen Data from the Tibetan Plateau

Caiming Shen,* Kam-biu Liu,† Lingyu Tang,‡ and Jonathan T. Overpeck§

*Atmospheric Sciences Research Center, State University of New York at Albany, and Nanjing Institute of Geology and Paleontology, Academia Sinica
†Department of Oceanography and Coastal Sciences, Louisiana State University
‡Nanjing Institute of Geology and Paleontology, Academia Sinica
§Institute for the Study of Planet Earth and Department of Geosciences, University of Arizona

This study applies a number of multivariate numerical techniques in the analysis of modern and fossil pollen data for the purpose of paleovegetational reconstruction in the Tibetan Plateau. Modern pollen spectra from 227 sampling sites in the Tibetan Plateau were classified into groups using cluster analysis and detrended correspondence analysis. By comparing the quantitatively derived groups with local vegetation types at sampling locations, the modern pollen spectra were assigned into five major vegetation types—shrubland, forest, meadow, steppe, and desert. These were used as a priori groups for discriminant analysis. A set of discriminant functions was derived that correctly classifies 93.3 percent of the surface samples. These functions were then applied to a fossil pollen record from Yidun Glacial Lake near the Sichuan–Tibet border to reconstruct the major pattern of vegetational changes since the last glacial maximum. The results show that steppe and meadow prevailed around the site during the Late Glacial period (17.3–11.5 ka BP). Regional vegetation changed from meadow to forest during the transitional period of 11.5–9.2 ka BP. Forests have persisted in the Yidun area since 9.2 ka BP. A comparison of our reconstructed paleovegetation with other proxy data suggests that the transition to forest around 9.2 ka BP followed the early Holocene intensification of the southwest Asian monsoon. Our study also demonstrates that discriminant analysis is a useful technique for reliable reconstruction of paleovegetation in the Tibetan Plateau. Key Words: discriminant analysis, modern analogue, pollen data, Tibetan Plateau, vegetation change.

Este estudio aplica un número de técnicas numéricas multivariadas en el análisis de datos de polen moderno y fósil con el propósito de la reconstrucción de la paleovegetación en la meseta tibetana. El espectro de polen moderno de 227 sitios de muestreo en la meseta tibetana se clasificó en grupos usando análisis de conglomerados y análisis de correspondencias. Al comparar los grupos derivados cuantitativamente con tipos de vegetación local en los lugares de las muestras, el espectro de polen moderno se clasificó en cinco grupos de vegetación importantes: matorrales, bosques, praderas, estepas y desierto. Estos se utilizaron como grupos a priori para el análisis discriminatorio. Se derivó un conjunto de funciones discriminatorias que clasifica correctamente un 93.3 por ciento de las muestras de superficie. Estas funciones se aplicaron luego a los registros de polen fósil del Yidum Glacial Lake cerca de la frontera entre Sichuan y Tibet para reconstruir el principal patrón de cambios vegetativos desde el último máximo glacial. Los resultados muestran que la estepa y la pradera dominaron el sitio durante el periodo del último glacial (hace 17.3-11.5 miles de años). La vegetación regional cambió de pradera a bosque durante el periodo transicional de hace 11.5 a 9.2 miles de años. Los bosques han persistido en el área.
Understanding the quantitative relationships between modern pollen rain and contemporary vegetation and climate is vital for the interpretation of vegetation and climate history based on fossil pollen data (Bent and Wright 1963; Davis 1967; McAndrews and Wright 1969; Webb and Bryson 1972; Webb et al. 1981; Overpeck, Webb, and Prentice 1985; Bartlein, Prentice, and Webb 1986; Delcourt, Delcourt, and Davidson 1987; Bush 1991; MacDonald and Edwards 1991). The identification of modern analogs, or the lack of them, is an important procedure for the quantitative reconstruction of past vegetation and climate. The mathematical techniques used to measure the degree of similarity between modern and fossil pollen spectra include discriminant analysis (K. B. Liu and Lam 1985; MacDonald and Ritchie 1986; MacDonald 1987; Sugden and Meadows 1989; K. B. Liu 1990; H. Y. Liu, Cuy, and Huang 2001; Horrocks and Ogden 2003; Oswald et al. 2003), principal components analysis (MacDonald and Ritchie 1986; MacDonald 1987), dissimilarity coefficients (Overpeck, Webb, and Prentice 1985; Huntley 1990; Gavin et al. 2003), and detrended correspondence analysis (DCA; Jacobson and Grimm 1986; Bush 1991). Although these numerical techniques have been applied to modern and fossil pollen data from a variety of vegetation types in different parts of the world (e.g., North America, Europe, South Africa), they have not been applied in the pollen and vegetation complexes of the Tibetan Plateau.

The vegetation of the Tibetan Plateau follows a distinct horizontal gradient controlled by the southwest Asian monsoon (Figure 1). Tropical rainforest and seasonal rainforest grow on the windward slopes of the Himalayas in the southernmost part of the plateau. The former consists of *Dipterocarpus*, *Artocarpus*, and *Shorea*, whereas the latter is dominated by a variety of evergreen oaks (*Cyclobalanopsis*, *Quercus*, *Lithocarpus*, *Castanopsis*), mixed with subalpine conifers such as *Picea*, *Abies*,

---

**Figure 1.** Map of the Tibetan Plateau showing the regional vegetation and the location of surface samples (● = moss polsters, ○ = topsoil, and ◇ = lake mud) and Yidun Glacial Lake.
Numerical Analysis of Modern and Fossil Pollen Data from the Tibetan Plateau 757

and Tsuga (Wu 1980; Tibetan Investigation Group 1988). Northward onto the Tibetan Plateau, the vegetation changes to subalpine conifer forest dominated by Abies, Picea, Pinus, Tsuga, and Sabina, and then to alpine meadow dominated by Kobresia, together with Rhododendron, Polygonum, Thalictrum, Anphalis, Potentilla, and Caragana. The driest regions in the west and north are characterized by high-cold steppe or arid steppe and desert dominated by Artemisia, Stipa, Ceroxoides, and Ephedra (Tibetan Investigation Group 1988). Due to the varied topography of the Tibetan Plateau, a complex pattern of altitudinal vegetation zonation occurs in each vegetation region. For example, montane conifer-hardwood mixed forest, montane conifer forest, subalpine shrub meadow, alpine meadow, alpine periglacial desert, and perpetual snow and ice can be found at different elevations between 2,500 and 6,000 m in the subalpine conifer forest region of the southeast. Therefore, a variety of local vegetation types including different forest, meadow, steppe, and desert communities can grow within the same vegetation region.

Two sets of investigations were undertaken in this article. First, modern pollen spectra from the Tibetan Plateau were classified into groups by means of cluster analysis and DCA. These pollen sample groups were interpreted in terms of modern vegetation based on field observations of the local vegetation occurring at the sampling sites. In this procedure, the pollen types that are statistically important for describing major vegetation types were also determined. In the second set of investigations, a set of functions was derived using discriminant analysis. These functions were used to further validate the classification of modern pollen spectra, and also to detect the modern analogs for the fossil pollen spectra from Yidun Glacial Lake in a case study for validating the applicability of discriminant functions to fossil pollen spectra.

Data and Methods

Data

The pollen data set used in the numerical analysis consists of 227 modern pollen spectra collected from moss polsters (179 samples), topsoil (27 samples), and lake bottom sediments (21 samples) in the Tibetan Plateau. Surface pollen samples were taken from a subalpine conifer forest region in the southeast, alpine steppe region in the south under the rain shadow of the Himalayas, an alpine meadow region in the east-central plateau, a high-cold steppe region in the central plateau, and dry steppe and arid desert regions in the northeastern plateau (Figure 1). Local vegetation types at the sampling sites include different forest, shrub, meadow, steppe, and desert communities. A complete list of the surface samples and their site characteristics can be found in Shen (2003).

For data standardization and consistency, twenty pollen types were included in the numerical analysis based on their quantitative abundance or their ecological importance. These twenty pollen types included major arboreal pollen types such as Abies, Picea, Pinus, Quercus, and Betula; shrub pollen types such as Rhododendron, Rosaceae, and Salix; and herb pollen types such as Poaceae, Asteraceae, Artemisia, Chenopodiaceae, and Cyperaceae. Minor but ecologically important pollen types include tree genera Tsuga and Corylus, as well as the herb taxa Ranunculaceae, Thalictrum, Caryophyllaceae, Polygonum, and Fabaceae. Pollen percentages for all surface samples were recalculated based on a sum of these twenty pollen types (Figure 2).

During fieldwork, the local vegetation composition around each sampling site was qualitatively described and recorded. Although quantitative data on plant abundance were not available for making a statistical comparison between pollen and vegetation data (e.g., Davis 1963; Parsons and Prentice 1981; Delcourt, Delcourt, and Davidson 1987), the descriptions of common plant taxa at the sampling sites provided valuable information for relating the pollen signatures to their local vegetation types. These local vegetation types were used here for interpreting the ecological significance of the modern pollen samples and their quantitatively derived groups.

Methods

Cluster analysis is a numerical technique that classifies samples into groups according to the dissimilarity between them. It has been widely used to classify modern pollen data (Birks 1973; Markgraf, D’Antoni, and Ager 1981; Dodson 1983; Lamb 1984). Birks and Gordon (1985) provided a detailed description of cluster analysis. In this study, Grimm’s (1987) CONISS, which uses the method of incremental sum of squares and Edwards and Cavalli-Sforza’s chord distance as dissimilarity coefficients, was applied to classify the non-sequential modern pollen samples into groups. Cluster analysis is used here as a numeric classifier for the objective classification of modern pollen data.

DCA is an ordination method commonly used to analyze modern and fossil pollen data (Jacobson and
Figure 2. Summary pollen percentage diagram for 227 surface samples and sample groups classified by cluster analysis. Pollen percentages were calculated based on a sum of twenty major pollen taxa.
Grimm 1986; Bush 1991). The results, expressed as a series of scores for the samples and pollen taxa, are plotted on the first two DCA axes to delineate the sample groups by clustering the close sample points and to reveal the behaviors of pollen taxa on those two axes. The objective of using DCA in this study was to cross-check the sample groups defined by the cluster analysis and determine the importance of different pollen types in different groups. Therefore, it is used here not only as a numeric classifier but also as a technique to reveal the structure of modern pollen data. DCA of surface samples was implemented by using data analysis of the TILIA software (Grimm 1990).

Discriminant analysis is a mathematical technique used to measure the degree of similarity between modern and fossil pollen assemblages (K. B. Liu and Lam 1985; MacDonald and Ritchie 1986; Sugden and Meadows 1989; Oswald et al. 2003). Unlike cluster analysis and DCA, the discriminant analysis is a statistical inference technique rather than classification method. It is used here not only to verify the classification conducted by cluster analysis and DCA, but also to determine how well the modern pollen assemblages can provide analogs for certain vegetation types, and to build the discriminant functions for identifying modern analogs from the fossil pollen spectra. The interpretation of results follows the procedures of K. B. Liu and Lam (1985). Software used in this procedure is SPSS 10.0 (SPSS, Inc. 1999).

Results

Cluster Analysis

The results of cluster analysis suggest that the 227 surface samples can be divided into fourteen groups (Figure 2). Groups TS1, TS2, and TS3 are dominated by Chenopodiaceae, Artemisia, and Poaceae, respectively. Groups TS4 and TS5 are both characterized by the highest percentages of Cyperaceae, but Poaceae pollen is more frequent in group TS4 than in group TS5. All samples in group TS6 have moderate amounts of Cyperaceae pollen, but this group is more heterogeneous than all previous groups because some samples have relatively high percentages of Artemisia and Chenopodiaceae pollen, whereas others have high percentages of Asteraceae pollen. There are no dominant pollen types in group TS7, but Cyperaceae, Artemisia, Poaceae, Asteraceae, Quercus, Betula, and Rosaceae are all fairly common. Group TS8 is dominated by Rhododendron pollen (40–90 percent). Groups TS9 through TS14 are characterized by high percentages of tree pollen such as Abies, Picea, Pinus, Quercus, and Betula. These six groups are distinguished from each other by variations in the pollen percentages of different tree taxa, Artemisia, and Cyperaceae.

A comparison between the groups of surface samples and their local vegetation types (Figure 3) shows that variations in modern pollen rain are related to differences in local vegetation. The group TS1, dominated by Chenopodiaceae pollen, reflects arid desert environments. Most samples of the group TS2 with high pollen percentages of Artemisia (>40 percent) are from dry steppe communities. Except for two samples, samples in group TS3 are from the Stipa steppe in the central plateau. Groups TS4 through TS7 are mainly from alpine or shrub meadows, and the high pollen percentages of Cyperaceae in these samples reflect the importance of sedges in these meadows, although higher percentages of Artemisia, Poaceae, and Asteraceae in some samples may reflect floristic variations among different meadow communities. All samples in group TS8 are from Rhododendron shrubland communities. Groups TS9 through TS14 are mainly composed of samples from forest, although some samples from shrubland, shrub meadow, meadow, and valley desert are also classified into these groups. The misclassification of surface samples from nonforest vegetation types into groups dominated by forest samples by cluster analysis is probably attributed to four causes. First, pollen from trees may be dispersed in areas above the treeline due to uphill aeolian transportation of pollen (Flenley 1979; Markgraf, D’Antoni, and Ager 1981). Second, some sampling sites are located in the ecotone between forest and nonforest communities. Third, pollen spectra are strongly affected by the local vegetation communities, especially individual trees growing at or near the sampling site. Finally, some sampling sites (e.g., five samples in group T10; see Figure 3) are situated in valley deserts surrounded by mountains where forests grow, so that the tree pollen derived from these montane forests may overwhelm the sparse pollen production from the local desert plants.

Detrended Correspondence Analysis

The geometric representation in the first two DCA axes accounts for 69 percent of the original variability and can thus be regarded as providing a reasonable summary of the information contained in the full twenty-dimensional space. The ordination of samples (sample scores) is directly related to the ordination of pollen types (variable scores). Samples having high scores on an axis are dominated by pollen types having high variable
Figure 3. Local vegetation types of the 227 surface samples in the fourteen groups classified by cluster analysis, and major vegetation types defined by comparison of local vegetation types and classification of surface samples. Dots indicate the local vegetation types of surface samples.
Figure 4. Detrended correspondence analysis (DCA) results of the surface samples: Sample scores on the first four DCA axes. Dots indicate the local vegetation types of surface samples.
Figure 5. Ordination diagram of detrended correspondence analysis (DCA) on the first two DCA axes.

The results of DCA appear to support the classification derived from cluster analysis. The assessment of the contribution of pollen types in the composition of samples from different vegetation types is consistent with that revealed by the pollen diagram. However, the samples from shrubland and shrub meadow cannot be classified clearly by either cluster analysis or DCA. Except for group TS8, which is composed solely of some samples from *Rhododendron* shrub, no other shrubland groups were identified by these two numerical techniques. Most of the shrubland samples were classified into forest groups.

**Discriminant Analysis**

Discriminant analysis requires an a priori classification of samples into groups. The conventional practice is that surface samples are classified into the required a priori groups based on the major vegetation regions in which the samples are located (K. B. Liu and Lam 1985; K. B. Liu 1990). In the Tibetan Plateau, especially

scores on that axis (Jacobson and Grimm 1986). The first axis, on which Chenopodiaceae has the highest positive value and *Rhododendron* has the highest negative value, separates desert samples from *Rhododendron* shrub and forest samples (Figures 4 and 5). The first axis of the ordination also clearly separates arboreal taxa, such as *Picea*, *Pinus*, *Abies*, *Betula*, and *Quercus* from the nonforest taxa such as *Artemisia*, *Poaceae*, *Asteraceae*, and *Cyperaceae*. The second axis separates the steppe samples, especially sagebrush steppe, from meadow and shrub meadow samples. The scores of samples and pollen types on this axis indicate that *Artemisia* and *Cyperaceae* make the largest contribution to the composition of samples from steppe and meadow, respectively, although *Poaceae* and *Asteraceae* are important to samples taken from both steppe and meadow. Figures 4 and 5 also show a clear separation among samples from different forest types, especially between mixed forest and broadleaved forest. It seems that birch forest occurs closer to mixed forest or conifer forest than oak forest as indicated by their modern geographic distributions.
in the topographically complex southeastern part, it is
difficult to classify surface samples based on the broad
vegetation region because there are a variety of azonal
or local vegetation types from forest to steppe commu-
nities, including valley desert in the forest region. Here
we assigned the modern pollen spectra to five major
vegetation types (i.e., desert, steppe, meadow, shrub-
land, and forest) as a priori groups by comparing surface
pollen sample groups classified by cluster analysis and
DCA with local vegetation types (Figure 3). These five
major vegetation types are defined on the basis of local
vegetation communities at the sampling sites of surface
samples in classified groups. They are not the same as
major vegetation regions in concept because samples in
the same major vegetation type may be located in differ-
ent vegetation regions. For example, spruce forest can be
found not only in the subalpine conifer forest region of
the southeastern plateau, but also in the alpine meadow
and shrubland region of the east-central plateau.

Five surface pollen sample groups (TS1, TS2–TS3,
TS4–TS7, TS8, and TS9–TS14) are interpreted to cor-
respond to the five vegetation types, respectively. One
sample from meadow was deleted from group TS1 be-
cause it contains an anomalously high percentage of
Chenopodiaceae pollen—probably distorted by the in-
corporation of an anther from a chenopod plant growing
at the sampling site. For the same reason, another nine
samples were deleted from the modern data set, includ-
ing three samples with anomalously high percentages
of Artemisia and Cyperaceae pollen from Rhododendron
shrub, rosaceous shrub, and steppe in group TS2, two
samples with anomalously high percentages of Cyper-
aceae pollen from Sabina forest and Rhododendron shrub
in group TS5, two samples with anomalously high per-
centages of Artemisia and Asteraceae pollen from ros-
aceous shrub and shrub meadow in group TS6, and two
samples with anomalously high percentages of Artemisia
pollen from pine oak forest and rosaceous shrub in group
TS10. In addition, five samples from valley desert and
two samples from willow scrub in group TS13 were also
excluded from the a priori group of forest. Ultimately,
210 pollen spectra assigned to five a priori groups (Table
1) were used in the discriminant analysis.

The five group centroids on the first two discrim-
inant functions (Figure 6) are clearly separated from
each other. The samples of desert and shrubland are
most distinct from those of the other major vegeta-
tion types. However, some overlapping occurs between
steppe, meadow, and forest samples. Some samples are
from the ecotonal areas between steppe and meadow or
between meadow and forest, so this overlapping reflects
the transitional character of the vegetation and their pa-
lynological signatures. This character is also indicated
by the classification results of discriminant analysis
(Table 2) and by the “probability of modern analog”
(Figure 7), which is a measure of similarity between a
sample and all other samples classified within the
same group. It is determined by the chi-square distance
between a sample and its assigned group centroid de-
fined by discriminant functions (see K. B. Liu and Lam
1985 for more detailed explanation). A comparison
of the predicted group memberships with the a priori
groups shows that 93.3 percent of all the samples are
correctly classified. The misclassified samples generally
have very low values (<0.2) of probability of modern
analog, indicating that they are away from the group
centroid; that is, they are from the transitional areas or
they are unique among members of their own groups
(K. B. Liu and Lam 1985). High probability of modern
analog (>0.5) for most surface samples suggests that
these samples can be regarded as typical of their groups
(i.e., the major vegetation types). Typically, transitional
and unique samples are also detected by the vegeta-
tional zonal index (VZI; Figure 7), an index converted
from probabilities of group membership between the
predicted and the second most probable groups (K. B.
Liu and Lam 1985).

In summary, the major vegetation types are well de-
finite by cluster analysis, DCA, and discriminant anal-
ysis. Most of the modern pollen samples are typical
for these major vegetation types as reflected by fairly
high values of probability of modern analogue. These
numerical techniques are also useful in identifying the
dominant pollen types representing each major vege-
tation type and in depicting the transitional or unique
samples. Thus, the discriminant functions can be used
to identify modern analogs for fossil pollen spectra.

A Case Study in Paleovegetational Reconstruction
Using Discriminant Analysis

The fossil pollen diagram selected in this case study
was derived from Yidun Glacial Lake (30°17.95′N,
99°33.08′E). The lake is situated in the subalpine
conifer forest region of the southeastern Tibetan
Plateau (Figure 1). At 4,470 m above sea level, the
lake is locally surrounded by alpine shrub meadow domi-
nated by Cyperaceae and Rhododendron with some Salix.
Treeline, formed by Picea, occurs at about 4,400 m and
just 2 km south of the lake. Dense forests consisting of
spruce, fir, pine, birch, oak, and juniper occur at lower
elevations down the valley.
A 4.5-m core was taken from 2.1 m of water. The sediment stratigraphy of this core consists of 2.8 m of clayey gyttja (organic matter content 10–20 percent) overlying clay. Three accelerator mass spectrometry (AMS) radiocarbon dates were obtained from organic sediment (Figure 8). A date of 2,070 ± 60 14C years BP was obtained from a bulk sediment sample at the core top (0–1 cm), indicating that the dates must be corrected for the hard-water effect. This observation is confirmed by another AMS date of 2,040 ± 50 14C years BP obtained from a modern sample of aquatic sedge taken from the lake edge. Another two dates of 5,710 ± 70 and 11,330 ± 140 14C years BP were obtained from clayey gyttja at a depth of 142 cm and 275 cm, respectively.

Table 1. Summary statistics of modern pollen rain in different major vegetation types

<table>
<thead>
<tr>
<th>Pollen taxa</th>
<th>Shrubland</th>
<th>Forest</th>
<th>Meadow</th>
<th>Steppe</th>
<th>Desert</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M  SD</td>
<td>M  SD</td>
<td>M  SD</td>
<td>M  SD</td>
<td>M  SD</td>
</tr>
<tr>
<td>Abies</td>
<td>0.1 0.3</td>
<td>5.3 5.4</td>
<td>0.1 0.4</td>
<td>0.1 0.2</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Pinus 1.9 0.9</td>
<td>15.2 13.3</td>
<td>1.7 3.2</td>
<td>0.4 1.2</td>
<td>0.0 0.1</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Tsuga 0.0 0.1</td>
<td>0.2 0.6</td>
<td>0.0 0.1</td>
<td>0.0 0.1</td>
<td>0.0 0.1</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Quercus 3.2 2.4</td>
<td>16.9 17.4</td>
<td>2.3 3.7</td>
<td>0.1 0.1</td>
<td>0.0 0.0</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Betula 4.0 2.3</td>
<td>16.9 12.1</td>
<td>2.1 2.5</td>
<td>0.4 0.1</td>
<td>0.1 0.1</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Corylus 0.0 0.1</td>
<td>1.1 2.8</td>
<td>0.3 1.2</td>
<td>0.1 0.2</td>
<td>0.0 0.0</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Salix 0.4 0.5</td>
<td>2.8 4.5</td>
<td>1.0 2.5</td>
<td>0.3 0.1</td>
<td>0.1 0.3</td>
<td>0.4 1.2</td>
</tr>
<tr>
<td>Rhododendron 64.4 13.2</td>
<td>3.1 6.4</td>
<td>0.9 2.8</td>
<td>0.0 0.0</td>
<td>0.3 1.2</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Rosaceae 0.0 0.0</td>
<td>6.6 6.7</td>
<td>2.9 4.1</td>
<td>0.4 0.8</td>
<td>1.3 4.9</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Poaceae 4.1 1.7</td>
<td>3.3 3.9</td>
<td>7.3 9.1</td>
<td>28.0 20.3</td>
<td>1.1 1.7</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Asteraceae 3.6 2.0</td>
<td>1.9 2.1</td>
<td>4.2 4.9</td>
<td>4.4 5.0</td>
<td>0.8 1.8</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Artemisia 3.5 3.8</td>
<td>7.8 7.5</td>
<td>8.3 10.3</td>
<td>23.6 33.0</td>
<td>5.9 9.1</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Chenopodiaceae 0.1 0.2</td>
<td>1.0 2.7</td>
<td>3.8 5.5</td>
<td>13.3 11.8</td>
<td>81.3 15.8</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Rumex 0.2 0.3</td>
<td>2.1 2.5</td>
<td>1.7 3.0</td>
<td>2.3 2.1</td>
<td>0.3 0.5</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Thalictrum 0.2 0.3</td>
<td>1.3 2.3</td>
<td>2.3 4.0</td>
<td>0.9 1.9</td>
<td>0.9 2.4</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Caryophyllaceae 0.6 0.7</td>
<td>0.6 1.3</td>
<td>1.5 3.6</td>
<td>1.1 1.5</td>
<td>0.4 1.6</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Polygonum 8.3 14.2</td>
<td>1.7 3.0</td>
<td>1.4 3.5</td>
<td>0.6 1.1</td>
<td>0.1 0.3</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Fabaceae 0.0 0.1</td>
<td>1.6 2.8</td>
<td>1.3 2.0</td>
<td>2.1 2.4</td>
<td>0.4 1.2</td>
<td>0.0 0.0</td>
</tr>
<tr>
<td>Cyperaceae 5.2 8.0</td>
<td>7.2 11.0</td>
<td>56.4 23.2</td>
<td>19.1 15.3</td>
<td>6.7 12.2</td>
<td>0.0 0.0</td>
</tr>
</tbody>
</table>

Table 2. Classification results of the surface samples by discriminant analysis

<table>
<thead>
<tr>
<th>Actual group</th>
<th>No. of samples</th>
<th>Shrubland</th>
<th>Forest</th>
<th>Meadow</th>
<th>Steppe</th>
<th>Desert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shrubland</td>
<td>7</td>
<td>7 (100)%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Forest</td>
<td>75</td>
<td>0</td>
<td>72</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Meadow</td>
<td>84</td>
<td>0</td>
<td>1</td>
<td>74</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Steppe</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>Desert</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>

Total number of samples: 210
Number of misclassified samples: 14
Percentage of samples correctly classified: 93.3%

*aNumber of samples classified as that group.
*bPercentage of samples classified in that group.
respectively. The sediments of these two dating samples are similar to that of the top sample. As a first approximation, the \(^{14}\text{C}\) ages of these dated horizons were corrected by subtracting 2,070 \(^{14}\text{C}\) years from these dates. The age of the core is thus estimated by extrapolation and interpolation of the corrected dates. Similar problems with radiocarbon dating attributable to the hard-water effect have been reported from other lake-sediment studies in the Tibetan Plateau (e.g., Fontes et al. 1993; Morrill et al. 2006). The dating problem introduces some degree of uncertainty in the age model, but based on the relatively homogeneous nature of the clayey gyttja there is no reason to expect that the correction factor should vary significantly at least during the Holocene.

The size of Yidun Glacial Lake (about 200 m × 300 m) suggests that the major pollen source of the lake would be derived from regional rather than local vegetation (Jacobson and Bradshaw 1981). It is confirmed by the pollen sample from the top of the core, which is dominated by arboreal taxa such as \textit{Quercus}, \textit{Pinus}, \textit{Picea}, \textit{Abies}, and \textit{Betula}. It reflects forest (the regional vegetation in this area) rather than meadow (the local vegetation near the lake), as suggested by the results of discriminant analysis. Therefore, the fossil pollen spectra from Yidun Glacial Lake provide a history of regional vegetation changes in this area during the Late Glacial and Holocene.

Results of discriminant analysis on the fossil pollen spectra from Yidun Glacial Lake (Figure 8) show that most of the fossil pollen samples have high probabilities of modern analog (>0.5) except for some samples near the boundaries of pollen zones. The boundaries of pollen zones reflect the phases of vegetation change from one type to another; thus, the low probabilities of modern analog associated with these samples reflect their transitional character during episodes of major vegetation change. The lowest value of probability (close to zero) occurs at the basal sample, indicating the lack of modern analog for this sample. This basal sample has relatively high arboreal pollen percentages but very low total pollen concentration value, suggesting that the arboreal pollen grains were probably derived from long-distance transport.
Late Glacial (17.3–11.5 ka; Pollen Zones YGL 5 and 6). In the initial period of lake formation, the vegetation was dominated by steppe, as suggested by one sample with a VZI of 4.0 in zone 6. From 16.3 to 13.5 ka BP, the Yidun area was occupied by subalpine meadow (VZI \(\approx 3.0\)) consisting of Cyperaceae, Artemisia, Poaceae, and Caryophyllaceae. However, forests probably began to appear after 14.5 ka BP at lower elevations down the valleys.

Late Glacial to Early Holocene (11.5–9.2 ka; Pollen Zone YGL4). This interval includes the transition from meadow to forest (VZI \(\approx 2.0\)). The transition is marked by an increase in arboreal pollen and a dramatic decline in Cyperaceae. Forests invaded the Yidun area and started to move up the valley. The meadow was finally replaced by forests around 9.2 ka BP. The slightly lower values of the probability of modern analog for samples in this interval also indicate the transitional character of the vegetation.

Early Holocene to Middle Holocene (9.2–6.8 ka BP; Pollen Zone YGL3). Pollen assemblages of this interval are dominated by the arboreal taxa Betula and Pinus. Abies, Picea, and Quercus pollen is also frequently present. Pollen concentration also reached the highest values of the whole sequence. As indicated by VZI values of \(\approx 2.0\), forest, probably a mixed coniferous and
Figure 8. Pollen diagram from Yidun Glacial Lake and paleovegetation reconstructed according to the vegetation zonal index and probability of modern analog (left).
deciduous broadleaved forest, occurred in this area during this interval.

**Middle Holocene to Late Holocene (6.8–2.5 ka BP; Pollen Zone YGL2).** Pollen assemblages are characterized by high percentages of Pinus, continuous rise of Quercus percentages, and steady drop of Betula percentages. Birch, the dominant forest tree in the preceding period, was replaced by pine and oak. High values of the probability of modern analog in this period indicate the existence of forests that resemble typical modern ones.

**Late Holocene (2.5 ka BP to Present; Pollen Zone YGL1).** In this interval, Pinus pollen decreases, and Quercus replaces Pinus to become the most abundant pollen type, indicating an expansion of oak in the forest surrounding the lake. The results of discriminant analysis indicate that forests continued to dominate this area. The lower and more fluctuating values of the probability of modern analog suggest that the forest communities in this period are less stable than in the preceding period.

**Discussion**

**Numerical Analysis of Tibetan Pollen Surface Samples**

The source areas for modern pollen samples collected from lakes, moss polsters, and soils are known to differ, with lakes collecting pollen from a greater area than moss polsters or soils (Minckley and Whitlock 2000). Because modern pollen samples are used to interpret the fossil pollen records, most of which are from lake sediments, pollen from the mud–water interface of lakes is more desirable than pollen from moss polsters and soils (Jacobson and Bradshaw 1981). However, access to lakes is very limited due to complex topography and poor road conditions in the Tibetan Plateau, and moss polsters are rare in the drier and colder part of the plateau, so sampling from topsoil and thin mashes is often the only option. Our study reveals that the modern vegetation types are well represented by our suite of surface samples collected from lakes, moss polsters, and topsoil. Comparison of pollen samples derived from lake sediments and those collected from moss polsters or topsoil shows that they have broadly similar pollen assemblages, particularly in their percentages of nonarboreal pollen types (see Figure 2; Shen 2003). Such similarity is not surprising, as landscapes covered by meadows, steppes, and deserts are relatively flat and open in the Tibetan Plateau, resulting in a well-mixed and regionally homogenized pollen rain. On the other hand, the character of the local and extralocal vegetation communities around the sampling site is retained and reflected in the moss and topsoil samples. This is especially true for the forest and shrubland samples, in which the different dominant tree or shrub taxa in the floristic composition are clearly registered in the modern pollen rain. When we developed discriminant functions identifying the modern analogs of fossil pollen samples, we grouped the surface samples into five a priori major vegetation types based on the numerical analysis of pollen percentage data. For example, samples from different local forest communities (e.g., pine forest, oak forest, birch forest) were lumped into the forest group, so these major groups include both the local and regional pollen signatures. Moreover, the discriminant functions measure the distance between the fossil samples and the centroids of the groups to classify the fossil samples, implying a comparison of regional pollen signatures between the fossil samples from lake cores and the surface pollen samples. Therefore, the discriminant functions developed using the network of modern surface samples in the Tibetan Plateau can be used to reconstruct the paleo-vegetation. Other techniques used in the quantitative interpretation of fossil pollen spectra, such as DCA and modern analog method, require that modern pollen samples should be from the same type of sedimentary environment (Overpeck, Webb, and Prentice 1985; Jacobson and Grimm 1986; Birks 1995). With respect to the limitation of our modern pollen samples as mentioned earlier, the discriminant analysis would be a more suitable numerical technique than others in the Tibetan Plateau. Our case study from Yidun Glacial Lake demonstrates the feasibility of this approach.

**Pollen-Based Paleovegetational Reconstruction for the Tibetan Plateau**

The Tibetan Plateau is the least anthropologically disturbed region in China (Zhang et al. 2003). For this reason, compared with other more densely populated regions, the Tibetan Plateau is a place where the modern pollen rain is expected to represent more closely the “natural” vegetation, and the modern pollen–vegetation relationship derived from modern pollen rain studies can provide a more realistic basis for paleo-vegetational reconstruction using a modern analog approach. Thus, the Tibetan Plateau offers a suitable testing ground for applying numerical techniques for quantitative paleo-vegetational reconstruction by objectively linking modern pollen data sets with their fossil
counterparts. In this article, we use a case study from Yidun Glacial Lake to demonstrate the utility of using discriminant analysis on modern and fossil pollen data to reconstruct the history of vegetation changes in the eastern Tibetan Plateau during the Late Glacial and Holocene times.

The results of discriminant analysis show that the regional vegetation in the eastern Tibetan Plateau from 17.3 to 9.2 ka BP was steppe and meadow. It changed to forest after 9.2 ka BP and has persisted there since then. The vegetation transition from meadow to forest, which occurred from 11.5 to 9.2 ka BP, is clearly depicted by both the vegetation zonal index and probability of modern analog. Our findings can be compared to results from earlier investigations in the Tibetan Plateau. An 11,000-year pollen record from Lake Shayema in the southeastern plateau indicates that vegetation changed from deciduous broadleaved forest to mixed mesophytic forest around 9.1 ka BP (Jarvis 1993). Several pollen records spanning the Late Glacial and Holocene in the east-central plateau show that the area was desert steppe and alpine meadow before 9.2 ka BP, replaced by conifer forests after 9.2 ka BP (Shen, Tang, and Wang 1996; Yan et al. 1999). A transition from desert to steppe is found at 9.3 ka BP in the pollen record from a large lake (Selin Co) of the central plateau (Sun et al. 1993). A number of studies have revealed that a marked increase in strength of the southwest summer monsoon occurred around 9.2 ka BP (e.g., Overpeck et al. 1996; Fleitmann et al. 2003; Gupta, Anderson, and Overpeck 2003). The transition of vegetation around 9.2 ka BP in our record is in phase with the major intensification of the southwest Asian monsoon.

Although discriminant analysis results suggest that forest vegetation has persisted in the study region throughout the last 9.2 ka, careful inspection of the pollen record reveals a more dynamic history of Holocene vegetation changes. A mixed coniferous and deciduous broadleaved forest dominated by birch occurred in the study area from 9.2 to 6.8 ka BP, when fir-spruce forest occurred in the east-central plateau (Yan et al. 1999) and mixed mesophytic forest occurred in the southeastern plateau (Jarvis 1993). Meanwhile, steppe dominated in the northwestern and central plateau (Sun et al., 1993; Van Campo and Gasse 1993). Both our pollen record and the others indicate strong summer monsoon during this period. Between 6.8 and 2.5 ka BP, the study area was still occupied by mixed coniferous and deciduous broadleaved forest, although pine and oak gradually replaced birch as the dominant components of the forest.

Lower and fluctuating values of probability of modern analog after 2.5 ka BP suggest that the forest has become less stable during the Late Holocene. After 2.5 ka BP, pine declined in abundance and the forest was dominated by oak. Reconstructed precipitation shows that annual mean precipitation at the interval from 6.8 to 2.5 ka BP was higher than the present, and that it decreased to the present level after 2.5 ka BP in the study area (Shen et al. 2006). In the southeastern plateau, more evergreen sclerophyllous taxa were present from about 5.0 to 4.0 ka BP at Lake Shayema. After about 4.5 ka BP, components of the sclerophyllous evergreen broadleaved forest, including Lithocarpus/Castanopsis, Ericaceae, Aralia, Ilex, and Viburnum, dominated the vegetation, and continued to be predominant until about 1.0 ka BP, indicating increased temperature in winter and decreased precipitation in spring and early summer, and thus a weakened summer monsoon (Jarvis 1993). In the east-central plateau, a major shrinkage of conifer forest took place at 4.0 ka BP, and forests were replaced by meadows dominated by Cyperaceae thereafter (Yan et al. 1999). This fact suggests a decline in precipitation in this region after 4.0 ka BP. In the northeastern and central plateau, more desert components were present after 6.0 ka BP, indicating a significant decrease in precipitation (Sun et al. 1993; Van Campo and Gasse 1993). The time-transgressive nature of the major decrease in monsoonal precipitation in these sites along a modern precipitation gradient probably reveals the retreat of the summer monsoon front from the northwest to the southeast during the Mid- to Late Holocene (K. B. Liu, Yao, and Thompson 1998).

Although paleoecological evidence from Lhasa suggests that human activities might have contributed to forest decline in southern Tibet for the past four millennia (Miehe et al. 2006), this is probably not the case in the more humid and topographically more rugged region of the eastern and southeastern plateau. The pollen record from Lake Shayema in the southeastern margin of the plateau shows increased human disturbance only after 1 ka BP (Jarvis 1993). There is little that can be ascribed to human influence in our pollen record from Yidun Glacial Lake.

Conclusions

This research applies a number of numerical analytical techniques to modern pollen data from the Tibetan Plateau, including cluster analysis, DCA, and
discriminant analysis. It also applies the discriminant functions to fossil pollen data from a small lake in the eastern Tibetan Plateau for the purpose of paleovegetational reconstruction. The main results of this study are summarized here.

1. Two hundred and twenty-seven surface pollen samples from the Tibetan Plateau were classified into fourteen groups using cluster analysis and DCA. These groups broadly correspond with various local vegetation types around the sampling sites.

2. By comparing surface pollen sample groups obtained from cluster analysis and DCA with the local vegetation of sampling locations, modern pollen spectra were classified into five major vegetation types—shrubland, forest, meadow, steppe, and desert. These five vegetation types were then used as a priori groups for the discriminant analysis. Seventeen anomalous samples were excluded from the modern pollen data set in this procedure.

3. A set of discriminant functions was then derived using discriminant analysis that correctly classified 93.3 percent of 210 surface samples. The misclassified samples were mainly from the transitional areas between major vegetation regions.

4. A case study, in which discriminant functions were applied to a fossil pollen record from Yidun Glacial Lake, demonstrates the use of discriminant analysis as a useful, rigorous, and objective mechanism for the description of vegetation types in both modern and fossil pollen data. It provides a quantitative measure of the similarity of cases in the fossil pollen data to the classes identified from the modern pollen data, and thus provides a mathematical treatment of the degree to which there are direct analogs in the modern and fossil vegetation.

5. In the case study, the results of discriminant analysis suggest at least four major paleovegetation phases:
   (1) VZI values of 4.0 to 3.0 indicate the dominance of steppe and meadow from 17.3 to 11.5 ka BP;
   (2) low values of probability of modern analog characterize the transition from meadow to forest during the Late Glacial to Early Holocene period (11.5–9.2 ka BP);
   (3) VZI values of 2.0 indicate prevalence of forest from 9.2 ka BP to the present;
   (4) lower and more fluctuating values of probability of modern analog reflect reduced stability in forest communities during the past 2.5 ka. The major vegetation change from meadow to forest around 9.2 ka BP followed the Early Holocene intensification of the southwest Asian monsoon.

Acknowledgments

This research was supported by grants from the U.S. National Science Foundation (NSF grants ATM-9410491, ATM-0081941), The Chinese National Science Foundation (grants No. 49371068 and 49871078), and dissertation research grants from the Geological Society of America (GSA), Association of American Geographers (AAG), and the Robert C. West Field Research Award (LSU Department of Geography and Anthropology). We thank Dr. Richard G. Baker, Dr. Basil Gomez, Dr. Richard Aspinall, and three anonymous reviewers for comments on the article.

References


Numerical Analysis of Modern and Fossil Pollen Data from the Tibetan Plateau

Tibetan lakes (Sumxi Co, Longmu Co) from 13,000 B.P. Quaternary Science Reviews 12:875–87.


SPSS, Inc. 1999. SPSS Base 10.0 for Windows user’s guide. Chicago, IL: SPSS, Inc.


Webb, T., III, and R. A. Bryson. 1972. Late- and postglacial climatic change in the northern Midwest, USA.


Correspondence: Atmospheric Sciences Research Center, State University of New York, Albany, NY 12203, e-mail: cshen@climate.cestm.albany.edu (Shen); Department of Oceanography and Coastal Sciences, School of the Coast and Environment, Louisiana State University, Baton Rouge, LA 70803, e-mail: kliu1@lsu.edu (Liu); Nanjing Institute of Geology and Paleontology, Academia Sinica, 39 East Beijing Road, Nanjing, 210008, China, e-mail: ly-tang@nigpas.ac.cn (Tang); Institute for the Study of Planet Earth and Department of Geosciences, University of Arizona, Tucson, AZ 85721, e-mail: jto@u.arizona.edu (Overpeck).