Quaternary Research **54,** 144–154 (2000)
doi:10.1006/qres.2000.2143, available online at http://www.idealibrary.com on IDEAL®

Probability That a Fossil Absent from a Sample Is Also Absent from the Paleolandscape

Robert S. Nowak[1]

Department of Environmental and Resource Sciences, Mail Stop 370, University of Nevada, Reno, Reno, Nevada 89557

and

Cheryl L. Nowak and Robin J. Tausch

USDA Forest Service, Rocky Mountain Research Station, 920 Valley Road, Reno, Nevada 89512

E-mail: nowak@scsr.nevada.edu

Received June 30, 1999

A procedure to calculate probability limits for the inference that the absence of a taxon from a woodrat (*Neotoma*) midden implies absence from the paleolandscape uses paired samples of modern vegetation communities and paired samples of paleocommunities. Assumptions are: (1) each member of a sample pair is an independent measure of the same vegetation assemblage; (2) behavioral patterns of woodrats are the same as each midden in a paired sample is constructed; and (3) the probability of fossilization is zero when a taxon is absent from the vegetation. The procedure provides a logical test of data consistency: the upper probability limit of making a false inference should be greater than the lower limit. Averaged over 140 plant taxa, the upper and lower probability limits for a false inference were 11 and 7%, respectively. More than 70% of taxa passed the logical test, indicating a reliable procedure. For many taxa that failed the logical test, four potential explanations account for this failure, two of which can be solved by simply increasing sample sizes. Using analogous assumptions, the procedures are applicable to other types of stratigraphic sampling such as macrofossils from sediment cores or fossils from biostratigraphic units. © 2000 University of Washington.

Key Words: **biostratigraphy; woodrat (*Neotoma*) middens; paleovegetation; missing fossils; Great Basin vegetation.**

INTRODUCTION

When a particular fossil is found in a properly sampled and dated stratum, the fundamental tenet of biostratigraphy is that the organism was located within the area and was present during the time period represented by that stratum. However, when fossils of an organism are not found in a sample, we are not sure if that organism was truly absent from that time or locale or if that organism simply was not preserved. The ability to differentiate true absence from the vagaries of fossilization and sampling is critical for interpretation of the fossil record (Holland 1995), such as when modeling biotic responses to climate change (e.g., Grayson and Livingston, 1993; Beck, 1996) or testing evolutionary theories (e.g., Boucot, 1996; Brett *et al.*, 1996; McKinney *et al.*, 1996). Statistical techniques can assess the completeness of the fossil record (Foote and Sepkoski, 1999), determine endpoints of stratigraphic ranges (Marshall, 1994; Solow, 1996), and define community assemblages of taxa (Bennington and Bambach, 1996), but these approaches generally require large data sets with accurate stratigraphic ranges to be used effectively.

In this paper, we describe a procedure to calculate the probability that the absence of a taxon from a stratigraphic sample provides reliable information about its absence from the paleolandscape. The concept is developed based upon plant macrofossils recovered from woodrat (*Neotoma*) middens in western North America and respective assumptions related to the construction of the middens. However, with broader definitions of terms and analogous assumptions, the procedure potentially is applicable to other paleobiological data sets, such as macrofossils from sediment cores or fossils from biostratigraphic units. We demonstrate the utility of this procedure by estimating the probabilities of making correct (true) and incorrect (false) inferences about the absence of more than 140 plant taxa from the paleolandscape.

METHODS, ASSUMPTIONS, AND MATERIALS

Four possible states exist concerning the presence or absence of a taxon in a stratigraphic sample and in the contemporaneous vegetation:

(1) *true presence,* denoted as (1,1), present in the stratum and present in the surrounding vegetation;

Supplementary data for this article may be found on the journal home page (http://www.academicpress.com/qr).

[1] To whom correspondence should be addressed. E-mail: nowak@scsr.nevada.edu.

0033-5894/00 $35.00
Copyright © 2000 by the University of Washington.
All rights of reproduction in any form reserved.

(2) *false presence,* denoted as (1,0), present in the stratum but absent from the surrounding vegetation;

(3) *false absence,* denoted as (0,1), absent from the stratum but present in the surrounding vegetation;

(4) *true absence,* denoted as (0,0), absent from the stratum and absent from the surrounding vegetation.

(Note that we use the term "vegetation" in this paper to indicate the plants within the specific geographic area around the midden from which the fossils in a midden sample are derived.) Although our concerns, as stated above, involve how to interpret absence from a stratum, the concerns from the perspective of probabilistic support are really the probability of making the correct inference (Popper, 1991). A correct or true inference is made for the first and fourth states, i.e., for true presence and for true absence. Thus, the probability of a true inference, denoted as $p(\mathrm{T})$, is the sum of the probabilities for states (1,1) and (0,0), which are denoted $p(1,1)$ and $p(0,0)$, respectively:

$$p(\mathrm{T}) = p(1,1) + p(0,0). \tag{1}$$

The false inference is made for the states of false presence and false absence; thus, the probability of a false inference, $p(\mathrm{F})$, is the sum of $p(1,0)$ and $p(0,1)$:

$$p(\mathrm{F}) = p(1,0) + p(0,1). \tag{2}$$

However, the theoretical value for false presence, $p(1,0)$, is zero. For a fossil to occur in the stratum when it is absent from the vegetation, we must invoke extraordinary mechanisms, i.e., unknown events or *ad hoc* hypotheses. In practice, we recognize that unknown events occur, but their probability is negligible (otherwise, we would know them). Although false presence in the vegetation can occur if proper sampling procedures are not followed (e.g., different strata are mixed together), this scenario represents a sampling error rather than a true probability that the state (1,0) occurs. Given that $p(1,0)$ is not detectably different from zero (i.e., unknown events are exceedingly rare), $p(\mathrm{F})$ reduces to

$$p(\mathrm{F}) = p(0,1). \tag{3}$$

As a robust test of these equations, two independent data sets were used that coetaneously estimated an upper limit and a lower limit for both $p(\mathrm{T})$ and $p(\mathrm{F})$.

The first data set consists of 27 paired samples of the modern vegetation assemblage. One sample of each pair consists of the taxa present in a recent midden sample (radiocarbon date $<$200 yr B.P. or an uncemented midden sample), and the other is from our survey of modern vegetation within a 100-m radius of the midden. The 100-m survey is based upon studies of woodrat foraging (Wells, 1983; Betancourt *et al.,* 1986). The underlying assumption for pairing samples is that each sample is an independent measure of the same vegetation assemblage. The frequency of each of the four possible states is computed for each individual taxon by summing across all paired samples then dividing by the total. If we assume that each frequency estimates the probability for each respective state, then $p(\mathrm{T})$ is computed from Eq. (1) and $p(\mathrm{F})$ is computed from Eq. (3). However, the *observed* frequency of false presence (i.e., present in the recent midden material but not observed in the vegetation) may be detectable for some species. Unknown events or *ad hoc* hypotheses do not need to be invoked for this scenario to occur. Rather, the false presence can be explained by differences in how woodrats and people sample the vegetation. Woodrats sample the vegetation both over a period of many years (Thompson, 1985, 1990) and potentially with a more complete exploration of the area. In contrast, our observations generally are made at one time with a more limited coverage of the potential source area. For example, a number of short-lived species were not observed near one of our previous midden locales during a series of drought years but were observed during a single wet year (Nowak *et al.,* 1994a). Thus, these spatial/temporal differences in vegetation sampling between woodrats and people may result in a detectable frequency of false presence. If we assume that these occurrences of false presence represent ambiguous cases (i.e., they may represent unknown events or they may represent human sampling errors), then Eq. (3) underestimates the value of $p(\mathrm{F})$ and represents the estimated lower limit of $p(\mathrm{F})$ (probability estimate 4 in Table 1). On the other hand, if all occurrences of the false presence were due to human sampling errors (i.e., if we had sampled better, we would have found the species in the vegetation), then $p(1,0)$ should be added to $p(1,1)$ before Eq. (1) is used with the paired samples of modern vegetation to calculate $p(\mathrm{T})$. This revision represents the estimated upper limit for $p(\mathrm{T})$ (1 in Table 1).

The second data set utilizes 35 paired samples of paleovegetation, where each pair of midden samples was from the same locale and had similar radiocarbon dates (i.e., the standard error bars associated with the radiocarbon dates overlap). As with the modern samples, the frequencies of four possible states are computed across all these pairs of fossil middens for each individual taxon and are used as probability estimates for each state. Before $p(\mathrm{T})$ and $p(\mathrm{F})$ are derived, two assumptions need to be explicitly stated: (1) each pair of fossil midden samples represents two independent samples of the same paleovegetation assemblage, and (2) behavioral patterns of the woodrats were similar as both middens were constructed. Given these assumptions, we expect that if a species is present in one sample of the pair, it should also be present in the other. Likewise, if the species is absent from one sample of the pair, it should be absent from the other. These states are directly analogous to true presence and true absence, and even though we do not have direct observations of the vegetation, the frequency of these states are estimates of $p(1,1)$ and $p(0,0)$. Unfortunately, cases where the taxon occurs in one member of the pair but not in the other are ambiguous. These ambiguities

TABLE 1
Equations to Estimate Probability for True and False Inferences[a]

Probability estimate	Source of data	Equation
1. p(T): upper limit	Recent midden/modern vegetation pairs	$p(1,1) + p(0,0) + p(1,0)$
2. p(T): lower limit	Midden/midden paleovegetation pairs	$p(1,1) + p(0,0)$
3. p(F): upper limit	Midden/midden paleovegetation pairs	$p(1,0) + p(0,1)$
4. p(F): lower limit	Recent midden/modern vegetation pairs	$p(0,1)$

[a] Sources of data and equations used to calculate probabilities of true, p(T), and of false, p(F), inferences for plant macrofossils retrieved from woodrat middens: $p(1,1)$ is the probability of an individual taxon being present in both samples that comprise a pair; $p(0,0)$ is the probability of absent in both samples; $p(1,0)$ is the probability of present in the first sample but absent in the second; and $p(0,1)$ is the probability of absent in the first sample but present in the second.

may represent either one of the false states (i.e., false presence or false absence) or errors due to false assumptions (e.g., the vegetation did change between the deposition of the strata and thus our assumption of the same vegetation is false, or large variations in woodrat behavior, such as feeding preferences, occurred). Because the sampling techniques are designed to minimize stratigraphic sampling errors (Spaulding *et al.*, 1990; Nowak *et al.*, 1994b), sampling error is assumed to be negligible. Thus, the estimate of p(T) computed from the fossil midden data using Eq. (1) underestimates the true value of p(T) because some of the ambiguous cases should have been state (1,1), and these sample pairs represent the estimated lower limit for p(T) (probability estimate 2 in Table 1). For the paired samples of paleovegetation, Eq. (2) must be used to compute p(F) and represents the estimated upper limit for p(F) (3 in Table 1).

Two issues need to be addressed concerning the underlying assumption that paired samples represent the same vegetation. First, we explicitly define "vegetation" for woodrat middens as those plants located within 100 m of the midden, with the unstated assumptions that woodrats forage that far but no further and that this value is the same among all animals. Although 100 m is reasonable given previous studies of woodrat foraging behavior (Wells, 1983; Betancourt *et al.*, 1986), we need to consider whether an error in this assumption leads to systematic biases in the results or simply increases random sampling error. Systematic bias invalidates our technique, whereas increased random error influences the sample size needed to get a valid result. Our method already takes into account the potential scenario where woodrats sample beyond 100 m and a plant only exists at distances greater than 100 m from the midden. This scenario results in a false presence, and as indicated above for the paired samples of modern vegetation, we treat these as human sampling error and use this information to estimate the lower limit for p(F). Nonetheless, this scenario is unlikely to exist because the odds of a particular plant species always lying just outside 100 m from the midden for the 10- to 100-year period of midden stratum construction at all of the midden locales is effectively zero. Similarly, the potential scenario that leads to a false absence (i.e., woodrats never forage to 100 m and a particular plant species always exists outside the foraging range but within 100 m of the midden) is also unlikely to occur. Interestingly, the problem of defining "vegetation" is less severe for the paired samples of paleovegetation because we simply need to assume that the geographic area sampled by woodrats was the same size for both midden samples of the pair. Thus, our definition of vegetation likely affects sampling error, but it does not lead to systematic error.

The second issue regarding paired samples representing the same vegetation concerns variations in vegetation throughout time caused either by natural processes or by impacts of historical land use. Because each member of a sample pair does not represent exactly the same time period, a directional change in vegetation potentially could result in a systematic error. Our technique already incorporates the scenario where a species has been lost due to historic land use changes, i.e., the state of false presence, where a plant is present in a modern midden but not in our modern vegetation survey. Furthermore, our technique also accounts for vegetation changes between each sample of paired paleovegetation middens, i.e., the states of false presence and false absence are summed to estimate the upper limit of p(F). Although the scenario of a false absence for paired modern samples, i.e., a species invaded the landscape after midden construction was completed, is a realistic concern, our results do not indicate a systematic error for all species (see Graminoid Herbs section below). A factor in our study that helps reduce the potential for systematic error is that the mean difference in age between paired samples is approximately 80 yr, which is within the decades-to-century time integration that occurs during construction of a midden stratum (Thompson, 1985, 1990). Hence, the time period represented by each sample of a pair overlapped for most paired samples, and a directional change in vegetation is more likely to increase sampling error than to introduce a systematic bias.

A list of sample locales, midden identification numbers, and radiocarbon ages is archived with Academic Press. In addition, the complete data sets for 4 taxa (*Amsinckia tessellata*, *Chaenactis douglasii*, *Cordylanthus ramosus*, and *Cordylanthus* spp.) are also archived with Academic Press to provide detailed examples of our data analyses.

RESULTS

Tree Taxa

Estimates of the upper and lower limits for p(T) and p(F) were computed for 20 tree taxa that were present in both modern and paleovegetation samples. Of these taxa, four species were relatively abundant in both data sets (i.e., occurred in at least 15% of the pairs in each data set; Table 2A). *Pinus monophylla* and *Juniperus osteosperma* occurred most frequently in the pairs. The occurrence over all four species and over both data sets averaged 39% of the sample pairs (Table 3A). Lower probability limits for a true inference were greater than 90% for these species. Upper probability limits for a false inference were less than 10% for all relatively abundant species and low (less than 3%) for two species (*Cercocarpus ledifolius* and *Juniperus osteosperma*). Averaged over all four tree species, the mean probability of a false inference was less than 5%. For these relatively abundant trees, the upper probability limit was greater than or equal to the lower limit, as would be expected from logic. This simple, logical test of the internal consistency of the results indicates that these probability estimates are reliable for all four species. Given that the upper limit for p(F) was between 0 and 10% for the trees and that the methodology provides reliable probability estimates, we conclude that the inference "absent from a particular woodrat midden implies absence from the paleovegetation" is false less than 10% of the time for these abundant tree species.

Eleven tree species occurred, on average, in only 7% of the paired vegetation samples (Tables 2B and 3B), and some plant specimens could only be identified to the genus or family level (Tables 2C and 3C). As with the more abundant tree species, the estimated probability of a false inference was less than 10% for all these relatively rare tree taxa. However, three species (*Juniperus occidentalis*, *Populus tremuloides*, and *Salix scouleriana*) failed the logical, internal consistency test that the upper probability limit is greater than or equal to the lower limit. Upon close inspection of the data, we noticed that one member of a pair may have plant specimens identifiable to the species level, but the other member would have specimens only identifiable to a higher taxonomic level. Thus, a potential explanation for why these three species failed the logical test was that the genus/family-level specimens actually were the respective species but could not be identified to the species level because they lacked sufficient diagnostic characteristics. To test this potential explanation, we recalculated the probability limits for these species assuming that the genus/family taxa represented specimens at the species level. For both *Juniperus occidentalis* and *Salix scouleriana*, the recalculated upper and lower limits for p(F) were 0.000; thus, both species now passed the logical test. For *Populus tremuloides*, the probability estimates did not change with the recalculations, and the logical test still failed for this species. Thus, limits in our ability to identify some plant specimens introduces some uncertainty in the procedure for at least some species.

Shrub Taxa

Thirteen shrub species were relatively abundant in the data and occurred on average in almost half the sample pairs (Tables 2D and 3D). Mean values of upper and lower limits of p(F) for these shrubs were greater than those for trees, and two species had a lower limit for a false inference over 20%. Averaged over all the abundant shrub species, the probability of a false inference was between 10 and 20%. Two species (*Holodiscus dumosus* and *Tetradymia glabrata*) failed the logical test. For *Holodiscus*, the upper and lower probability limits were close to each other, and a single change of a false absence to either a true absence or a true presence would be sufficient for *Holodiscus* to pass the logical test. Thus, the logical test for this species seemingly failed due to random chance, and additional sample pairs would likely resolve the probability estimates.

Thirteen shrub species were relatively rare in either one or both sets of samples (Table 2E). The probability limits for these species were greater than those of the abundant species, and only *Eriogonum microthecum* had a value for p(F) greater than 20%. However, only half of these species passed the logical test (Table 3E). For five of the six species that failed the logical test, their occurrence in paleovegetation samples was less than 10% and was also less than half that in modern samples. Thus, the probability estimates based on the modern vegetation sample pairs (i.e., p(T) upper limit and p(F) lower limit) are more reliable than those based on the paleovegetation sample pairs because of the greater sample size. Additional paleovegetation sample pairs are needed to make better estimates of the lower limit of p(T) and the upper limit of p(F). For the sixth species that failed the logical test (*Petrophyton caespitosum*), its occurrence in samples was less than 5%, and the upper and lower limits were so close in value that a single change between true and false inference would change the logical test from failure to pass. Because the logical test for *Petrophyton caespitosum* seemingly failed due to random chance, additional sample pairs again would resolve the probability estimates.

Four of the shrub genus/family taxa were relatively abundant in the vegetation samples (Table 2F), whereas three were relatively rare (Table 2G). (Note that although live plant specimens of the genera *Chrysothamnus* and *Ephedra* can be distinguished to the species level, plant specimens from middens cannot be readily differentiated; thus we pooled the species- and genus-level data.) *Chrysothamnus* and *Ephedra* occurred in at least one member of 70–100% of the sample pairs and had a probability of a false inference between 10 and 30%, which was similar to that for many of the most abundant shrub species. Only one of these seven taxa (*Epilobium* spp.) failed the logical test. Like many of the rare shrub species, *Epilobium* spp. occurred much more frequently in modern samples, suggesting that additional paleovegetation samples are needed to make more reliable estimates of the upper limit for p(F).

TABLE 2
Probability Estimates of True and False Inferences for Individual Taxa

			Probability estimates[b]				
	Frequency (%)[a]		*p* (true inference)		*p* (false inference)		
Growth form and relative abundance	Modern samples	Paleoveg. samples	Upper limit	Lower limit	Upper limit	Lower limit	Logical test[c]
A. Tree: abundant species							
Cercocarpus ledifolius	15	17	1.000	0.971	0.029	0.000	pass
Juniperus osteosperma	74	77	1.000	1.000	0.000	0.000	pass
Pinus monophylla	44	46	0.963	0.943	0.057	0.037	pass
Prunus virginiana	15	20	1.000	0.914	0.086	0.000	pass
B. Tree: rare species							
Abies concolor	7	6	1.000	1.000	0.000	0.000	pass
Alnus incana v. *tenuifolia*	7	6	1.000	1.000	0.000	0.000	pass
Cornus sericea	11	9	1.000	0.971	0.029	0.000	pass
Juniperus occidentalis	7	3	0.926	1.000	0.000	0.074	—
Pinus contorta	11	9	1.000	0.943	0.057	0.000	pass
Pinus jeffreyi	11	9	1.000	0.943	0.057	0.000	pass
Populus balsamifera	7	6	0.963	0.943	0.057	0.037	pass
Populus tremuloides	4	3	0.963	0.971	0.029	0.037	—
Salix exigua	7	6	1.000	1.000	0.000	0.000	pass
Salix scouleriana	7	3	0.926	0.971	0.029	0.074	—
Sambucus ceruleus	7	6	1.000	0.971	0.029	0.000	pass
C. Tree: genus/family taxa							
Juniperus spp.	4	3	1.000	1.000	0.000	0.000	pass
Pinaceae	15	11	1.000	0.914	0.086	0.000	pass
Populus spp.	7	6	0.963	0.943	0.057	0.037	pass
Salix spp.	7	14	1.000	0.971	0.029	0.000	pass
Sambucus spp.	7	9	1.000	0.971	0.029	0.000	pass
D. Shrub: abundant species							
Artemisia tridentata	100	66	0.778	0.743	0.257	0.222	pass
Atriplex confertifolia	59	54	1.000	0.829	0.171	0.000	pass
Brickellia microphylla	85	60	0.815	0.714	0.286	0.185	pass
Ericameria nanus	70	49	0.926	0.686	0.314	0.074	pass
Grayia spinosa	67	46	0.852	0.800	0.200	0.148	pass
Holodiscus dumosus	19	17	0.852	0.886	0.114	0.148	—
Leptodactylon pungens	30	29	0.926	0.857	0.143	0.074	pass
Purshia tridentata	37	37	1.000	0.886	0.114	0.000	pass
Ribes velutinum	41	31	0.926	0.800	0.200	0.074	pass
Rosa woodsii	41	46	0.889	0.743	0.257	0.111	pass
Salvia dorrii	26	37	0.963	0.914	0.086	0.037	pass
Symphoricarpos oreophilus	19	29	0.926	0.857	0.143	0.074	pass
Tetradymia glabrata	52	40	0.741	0.829	0.171	0.259	—
E. Shrub: rare species							
Amelanchier alnifolia	7	3	1.000	1.000	0.000	0.000	pass
Arceuthobium divaricatum	11	9	1.000	0.971	0.029	0.000	pass
Artemisia dracunculus	15	11	1.000	0.943	0.057	0.000	pass
Artemisia spinescens	22	6	0.963	1.000	0.000	0.037	—
Atriplex canescens	11	11	0.963	0.914	0.086	0.037	pass
Eriogonum heermannii	33	9	0.815	0.943	0.057	0.185	—
Eriogonum microthecum	26	9	0.741	0.914	0.086	0.259	—
Gnaphalium microcephalum	7	3	1.000	0.971	0.029	0.000	pass
Kochia americana	22	6	0.815	0.943	0.057	0.185	—
Petrophyton caespitosum	4	3	0.963	1.000	0.000	0.037	—
Prunus andersonii	11	20	1.000	0.914	0.086	0.000	pass
Symphoricarpos longiflorus	11	3	1.000	1.000	0.000	0.000	pass
Tetradymia canescens	22	9	0.889	0.943	0.057	0.111	—
F. Shrub: abundant genus/family taxa							
Artemisia (sec. Tridentatae)	100	100	1.000	0.943	0.057	0.000	pass
Asteraceae	37	66	1.000	0.571	0.429	0.000	pass

TABLE 2—*Continued*

Growth form and relative abundance	Frequency (%)[a]		Probability estimates[b]				Logical test[c]
			p (true inference)		*p* (false inference)		
	Modern samples	Paleoveg. samples	Upper limit	Lower limit	Upper limit	Lower limit	
Chrysothamnus {*viscidiflorus, nauseosus*}[d]	100	83	0.741	0.714	0.286	0.259	pass
Ephedra {*nevadensis, viridis*}[d]	70	69	0.889	0.686	0.314	0.111	pass
G. Shrub: rare genus/family taxa							
Epilobium spp.	19	6	0.852	0.943	0.057	0.148	—
Ribes spp.	4	6	1.000	0.943	0.057	0.000	pass
Rubus spp.	7	3	1.000	0.971	0.029	0.000	pass
H. Dicot herb: abundant species							
Amsinckia tessellata	70	66	0.852	0.771	0.229	0.148	pass
Chaenactis douglasii	41	26	0.852	0.857	0.143	0.148	—
Cryptantha pterocarya	19	29	1.000	0.886	0.114	0.000	pass
Cryptantha torreyana	19	31	1.000	0.714	0.286	0.000	pass
Mentzelia albicaulis	41	63	1.000	0.714	0.286	0.000	pass
Phacelia humilis	19	23	1.000	0.829	0.171	0.000	pass
I. Dicot herb: rare species							
Chaenactis stevoides	7	11	0.963	0.886	0.114	0.037	pass
Collinsia parviflora	15	14	1.000	0.943	0.057	0.000	pass
Cordylanthus ramosus	4	3	0.963	0.971	0.029	0.037	—
Cryptantha affinis	11	6	1.000	0.943	0.057	0.000	pass
Cryptantha circumscissa	15	11	0.889	0.886	0.114	0.111	pass
Cryptantha gracilis	7	26	1.000	0.771	0.229	0.000	pass
Cryptantha nevadensis	7	6	1.000	0.971	0.029	0.000	pass
Eatonella nivea	4	23	1.000	0.857	0.143	0.000	pass
Erodium cicutarium	44	14	0.963	0.914	0.086	0.037	pass
Galium aparine	7	29	0.963	0.829	0.171	0.037	pass
Glyptopleura marginata	4	3	1.000	0.971	0.029	0.000	pass
Lappula redowski	15	14	0.889	0.857	0.143	0.111	pass
Lithospermum ruderale	7	3	1.000	0.971	0.029	0.000	pass
Microseris lindleyi	4	3	1.000	0.971	0.029	0.000	pass
Mirabalis bigelovii	7	6	1.000	0.971	0.029	0.000	pass
Phacelia crenulata	22	14	1.000	0.914	0.086	0.000	pass
Phacelia glandlifera	7	11	1.000	0.914	0.086	0.000	pass
Phacelia incana	7	17	1.000	0.829	0.171	0.000	pass
Plagiobothrys kingii	11	26	1.000	0.800	0.200	0.000	pass
Polygonum douglasii	11	23	1.000	0.800	0.200	0.000	pass
Smilacina stellata	7	3	0.926	0.971	0.029	0.074	—
Stephanomeria spinosa	41	3	0.630	0.971	0.029	0.370	—
J. Dicot herb: abundant genus/family taxa							
Astragulus spp.	52	31	0.741	0.714	0.286	0.259	pass
Boraginaceae	26	34	1.000	0.743	0.257	0.000	pass
Castilleja spp.	15	31	1.000	0.771	0.229	0.000	pass
Chenopodium spp.	59	89	1.000	0.743	0.257	0.000	pass
Claytonia spp.	15	26	0.926	0.857	0.143	0.074	pass
Cryptantha spp.	33	23	0.778	0.800	0.200	0.222	—
Descurainia spp.	26	51	0.926	0.543	0.457	0.074	pass
Eriogonum spp.	37	43	0.815	0.686	0.314	0.185	pass
Lupinus spp.	44	63	1.000	0.571	0.429	0.000	pass
K. Dicot herb: rare genus/family taxa							
Allium spp.	15	6	0.889	0.943	0.057	0.111	—
Amaranthus spp.	7	11	0.926	0.886	0.114	0.074	pass
Arabis spp.	11	6	0.889	0.971	0.029	0.111	—
Arenaria spp.	11	20	1.000	0.857	0.143	0.000	pass
Aster spp.	11	17	1.000	0.857	0.143	0.000	pass
Balsamorhiza spp.	11	11	0.963	0.886	0.114	0.037	pass

TABLE 2—*Continued*

Growth form and relative abundance	Frequency (%)[a]		Probability estimates[b]				Logical test[c]
			p (true inference)		*p* (false inference)		
	Modern samples	Paleoveg. samples	Upper limit	Lower limit	Upper limit	Lower limit	
Brassicaceae	22	9	0.778	0.914	0.086	0.222	pass
Cactaceae	7	31	1.000	0.771	0.229	0.000	pass
Caryophyllaceae	4	3	0.963	0.971	0.029	0.037	—
Chamaesyce spp.	4	6	1.000	0.971	0.029	0.000	pass
Cirsium spp.	41	14	0.889	0.971	0.029	0.111	—
Crepsis spp.	11	11	0.963	0.914	0.086	0.037	pass
Fabaceae	7	9	1.000	0.914	0.086	0.000	pass
Galium spp.	4	14	1.000	0.886	0.114	0.000	pass
Gayophytum spp.	19	3	0.815	0.971	0.029	0.185	—
Geraniaceae	4	3	1.000	0.971	0.029	0.000	pass
Geranium spp.	7	6	0.963	0.971	0.029	0.037	—
Gilia spp.	26	14	0.852	0.886	0.114	0.148	—
Helianthus spp.	4	20	1.000	0.800	0.200	0.000	pass
Lepidium spp.	19	6	0.926	0.943	0.057	0.074	—
Lomatium spp.	19	3	0.815	0.971	0.029	0.185	—
Machaeranthera spp.	19	3	0.852	1.000	0.000	0.148	—
Malacothrix {*californica, sonchoides*}[d]	11	3	0.963	0.971	0.029	0.037	—
Nyctaginaceae	4	3	1.000	0.971	0.029	0.000	pass
Oenothera spp.	22	14	0.889	0.886	0.114	0.111	pass
Penstemon spp.	37	9	0.778	0.914	0.086	0.222	—
Phacelia spp.	15	14	0.926	0.914	0.086	0.074	pass
Polygonum spp.	4	14	1.000	0.857	0.143	0.000	pass
Scrophulariaceae	7	34	1.000	0.686	0.314	0.000	pass
Sphaeralcea spp.	7	6	0.963	0.943	0.057	0.037	pass
Stephanomeria spp.	4	3	0.963	0.971	0.029	0.037	—
Thelypodium spp.	26	3	0.741	0.971	0.029	0.259	—
Utica spp.	7	6	1.000	0.971	0.029	0.000	pass
Valerianaceae	7	3	1.000	0.971	0.029	0.000	pass
Viola spp.	4	9	1.000	0.943	0.057	0.000	pass
L. Graminoid herb: abundant species							
Bromus tectorum	93	20	0.704	0.914	0.086	0.296	—
Oryzopsis hymenoides	78	77	0.704	0.657	0.343	0.296	pass
Sitanion hystrix	96	57	0.333	0.629	0.371	0.667	—
Stipa comata	19	17	0.815	0.886	0.114	0.185	—
M. Graminoid herb: rare species							
Agropyron spicatum	15	14	0.889	0.914	0.086	0.111	—
Carex aurea	4	3	1.000	0.971	0.029	0.000	pass
Elymus cinereus	41	3	0.593	1.000	0.000	0.407	—
Poa sandbergii	70	11	0.370	0.943	0.057	0.630	—
Stipa speciosa	81	9	0.370	0.943	0.057	0.630	—
N. Graminoid herb: genus/family taxa							
Agropyron spp.	4	14	1.000	0.857	0.143	0.000	pass
Bromus spp.	7	6	1.000	0.943	0.057	0.000	pass
Carex spp.	22	9	0.852	0.943	0.057	0.148	—
Equisetum spp.	7	6	0.926	1.000	0.000	0.074	—
Poa spp.	26	17	0.852	0.914	0.086	0.148	—
Poaceae	96	80	1.000	0.743	0.257	0.000	pass
Stipa spp.	11	34	1.000	0.743	0.257	0.000	pass

[a] Percentage of paired samples that each taxon occurred in for 27 paired samples of modern vegetation and 35 paired samples of paleovegetation. Note that a taxon had to occur in only one member of a pair to be counted.

[b] Probabilities of true inference, *p*(T), and of false inference, *p*(F), calculated as described in Table 1.

[c] Simple, logical test that the upper probability limit is greater than or equal to the lower limit. Taxa that did not pass this test are indicated by dashes.

[d] Plant macrofossils cannot be differentiated to the species level.

TABLE 3
Mean Probability Estimates for Taxa Grouped by Relative Abundance and Growth Form

			Probability estimates[b]				
	Frequency (%)[a]		*p* (true inference)		*p* (false inference)		
Growth form and relative abundance	Modern samples	Paleoveg. samples	Upper limit	Lower limit	Upper limit	Lower limit	Logical test[c]
A. Tree: abundant species	37	40	0.991	0.957	0.043	0.009	100%
B. Tree: rare species	8	6	0.980	0.974	0.026	0.020	73%
C. Tree: genus/family taxa	8	9	0.993	0.960	0.040	0.007	100%
D. Shrub: abundant species	50	42	0.892	0.811	0.189	0.108	85%
E. Shrub: rare species	16	8	0.935	0.958	0.042	0.065	54%
F. Shrub: abundant genus/family taxa	77	80	0.908	0.728	0.272	0.092	100%
G. Shrub: rare genus/family taxa	10	5	0.951	0.952	0.048	0.049	67%
H. Dicot herb: abundant species	35	40	0.951	0.795	0.205	0.049	83%
I. Dicot herb: rare species	12	12	0.963	0.905	0.095	0.037	86%
J. Dicot herb: abundant genus/family taxa	34	43	0.910	0.714	0.286	0.090	89%
K. Dicot herb: rare genus/family taxa	11	10	0.944	0.920	0.080	0.056	66%
L. Graminoid herb: abundant species	72	43	0.639	0.772	0.229	0.361	25%
M. Graminoid herb: rare species	42	8	0.644	0.954	0.046	0.356	20%
N. Graminoid herb: genus/family taxa	25	24	0.947	0.878	0.122	0.053	57%

[a] Percentage of paired samples that taxa occurred in for modern vegetation and paleovegetation, averaged over all taxa within each growth form/relative abundance combination.
[b] Mean probabilities of true inference, p(T), and of false inference, p(F), over all taxa within each growth form/relative abundance combination.
[c] Percentage of taxa within each growth form/relative abundance combination that passed the logical test.

Dicot Herb Taxa

Six dicot herb species were relatively abundant, and twenty-two were relatively rare in the vegetation samples (Tables 2H and 2I). The lower limits of a false inference were similar between these two groups and averaged less than 5% (Tables 3H and 3I). However, the upper limit of a false inference for abundant dicot herbs averaged twice that of the rare species. Over 80% of the dicot herb species passed the logical test that the upper probability limit is greater than or equal to the lower limit. Of the four species that failed the test, three species (*Chaenactis douglasii*, *Cordylanthus ramosus*, and *Smilacina stellata*) had nearly identical upper and lower limits and thus likely failed the test due to random chance. The other species (*Stephanomeria spinosa*) was much more abundant in modern samples than in the paleovegetation samples. For all four species, additional samples would likely help resolve the probability estimates. The importance of additional sample pairs also is suggested by data for *Cordylanthus ramosus.* An additional 8% of the paleovegetation pairs had plant specimens that could only be differentiated to the genus level. If we assume that the *Cordylanthus* spp. specimens are actually *Cordylanthus ramosus* and pool these data, then the upper limit of p(F) increases to 0.086 and the taxon passes the logical test.

The probability of false inferences for abundant dicot herb genus/family taxa was relatively high and averaged between 10 and 30% (Table 3J), whereas that for relatively rare taxa averaged less than 10% (Table 3K). Five genera (*Cryptantha*, *Galium*, *Phacelia*, *Polygonum*, and *Stephanomeria*) also contained other plant specimens that were identified to the species level and occurred in both modern and paleovegetation samples. Data for these genus/species can be pooled in two ways: (1) assume that the genus-level specimens actually were a particular species and recalculate probability estimates for the individual species; (2) pool all individual species with the genus-level data and recalculate probability estimates for the genus. Interestingly, the first method only slightly narrowed the difference between upper and lower probability estimates for an individual species because genus-level specimens often were in the same sample as species-level specimens. The second method also generally narrowed the difference between upper and lower probability estimates, except for *Cryptantha* spp. For *Cryptantha* spp., the upper limit of p(F) increased to 0.343 and the lower limit decreased to 0.111, and the genus then passed the logical test. In addition, twelve genera had plant specimens that were identified to the species level in modern samples only (data not shown). For five genera (*Allium*, *Cirsium*, *Machaeranthera*, *Penstemon*, and *Utica*), combining the modern-only data with the genus-level data had little effect on the probability estimates for the genera. For six genera (*Balsamorhiza*, *Crepis*, *Descurainia*, *Gilia*, *Lepidium*,

and *Lupinus*), pooling the modern-only data with the genus-level data narrowed the difference between the upper and lower probability estimates. For the last genus (*Eriogonum*), including modern-only data greatly increased the lower limit of $p(\mathrm{F})$ to 0.593 and caused the logical test to fail.

Graminoid Herb Taxa

Graminoid herb species had the greatest lower limit of a false inference as well as the highest percentage of species that failed the logical test (Tables 2L, 2M, 3L, and 3M). At the genus/family level, probability estimates for graminoid herbs were similar to those of the other growth forms, but over 40% of the taxa still failed the logical test of data consistency (Table 3N). These greater probability estimates at higher taxonomic levels suggest that our inability to differentiate among species for the fossil material may be a major reason that so many graminoid herbs failed the logical test. Random chance and small sample sizes for paleovegetation also likely account for some data inconsistencies. Regardless of the reason, our ability to make inferences about the absence of graminoid herbs from the vegetation when they are absent from a midden sample is low for many graminoid taxa. For example, even for *Oryzopsis hymenoides,* which occurred in almost 80% of the sample pairs and passed the logical test, a false inference would be made over 30% of the time.

The occurrence of *Bromus tectorum* and the dicot herb *Erodium cicutarium* in "paleovegetation" samples initially seems incongruous, as these two species were introduced into North America by Europeans. However, all the midden samples that contained these species were less than 120 years old, which would be consistent with their recent and rapid invasions. For example, Mensing and Byrne (1998) documented that the range expansion of *E. cicutarium* near Santa Barbara, California, was so rapid that it actually preceded settlement by Spanish missionaries. These two taxa also provide insight about the potential for systematic errors between paired samples caused by directional changes in vegetation (see above). Both taxa potentially represent the scenario of a false absence for paired modern samples, i.e., species invaded the landscape after midden construction was completed. However, *E. cicutarium* passed the logical test and had a low probability of a false inference (Table 2I). Thus, systematic error from directional changes in vegetation did not occur in all cases where we might expect it and, hence, did not compromise our assumption that paired samples represent the same vegetation.

DISCUSSION

We have demonstrated a reliable method to estimate the upper and lower limits for the probability of false inference on the absence of an individual plant from the paleolandscape when that plant's fossils are absent from woodrat middens. Averaged over all taxa, the probability of a false inference is between 7 and 11%. For some taxa, such as *Juniperus osteosperma* and *Pinus monophylla*, the probability of a false inference is even lower, in these cases between 0 and 6%. For other taxa, especially graminoid herbs, the probability of a false inference can be higher, in some cases greater than 40%. Our method also provides a simple, logical test of internal data consistency. Seventy-two percent of the taxa passed this test, indicating a high degree of reliability for the probability estimates. The method is especially robust for species that are relatively abundant. For all abundant species except graminoid herbs, 87% passed the logical test.

This procedure is also generally reliable for relatively rare taxa (i.e., taxa that occurred in less than 15% of the sample pairs). Fewer rare species passed the logical test than the relatively abundant species (74% for all rare species except graminoid herbs versus 87% for abundant). However, the mean lower and upper limits of $p(\mathrm{F})$ for rare species (0.041 and 0.063, respectively) were *less* than those for abundant species (0.076 and 0.168), which suggests that the probability of making false inferences for rare species is lower than that for abundant species. The smaller estimates of $p(\mathrm{F})$ for rare species seems, at first, counterintuitive. Because a rare species in the vegetation is expected to have a low abundance in the respective midden strata, the portion of each stratum actually sampled is unlikely to contain the species. Thus, the state of true presence, (1,1), is expected to be at least as rare as the state of false absence, (0,1). On the other hand, the state of true absence, (0,0), should be the most abundant state for rare species. Because the probabilistic support for making a correct inference includes both true presence and true absence, the sum of a low value $p(1,1)$ plus a high value $p(0,0)$ yields a number much greater than $p(0,1)$. Thus, lower probabilities for rare species are rational.

For many of the taxa that failed the logical test of data consistency, we have presented three explanations for why they failed. First, some plant macrofossils lack diagnostic characteristics that allow identification to the species level. Consequently, we lacked definitive verification that the species was present in the second member of a paired sample, and the sample pair was incorrectly classified as false absence. Identification limitations appeared to be important for two tree species (*Juniperus occidentalis* and *Salix scouleriana*) and for a number of dicot and graminoid herb taxa. Pooling species-level and genus-level data often narrowed the difference between upper and lower probability limits and was sufficient for some taxa to pass the logical test. A second explanation for data discrepancies is random chance (i.e., the upper and lower probability limits are so close that a single change from false absence to true presence is sufficient for the taxon to pass the logical test). Random chance may be especially important for relatively rare taxa, as exemplified by the dicot herb *Cordylanthus ramosus.* As sample size increases, we expect that the upper and lower limits will converge toward the actual probability of a false inference. Thus, even though an abundant

taxon may fail the logical test because of random chance, the estimates of a false inference may still be reasonable approximations of p(F). Third, some taxa occurred much more frequently in the modern vegetation samples than in the paleovegetation samples. In these cases, we expect that probability estimates based on modern vegetation data are more reliable than those based on paleovegetation samples, and additional paleovegetation pairs are needed to resolve the upper limit for a false inference.

A fourth explanation for why taxa failed the logical test may be the selectivity of woodrats against particular taxa. Woodrat middens generally are accurate representations of species richness in the animal's environment (Dial and Czaplewski, 1990), but not necessarily of their abundance. When woodrats select against a certain taxon, then we expect a high occurrence of false absences in the paired samples of modern vegetation (i.e., present in the modern vegetation survey but absent from the recent midden sample) but a low occurrence of false absences in the paleovegetation samples (i.e., equally rare in each midden sample). Hence, the occurrence of the taxon in the modern samples should be much greater than in the paleovegetation samples. Furthermore, the lower limit of p(F), which is derived from the modern sample pairs, should be much greater than the upper limit p(F), which is derived from the paleovegetation pairs. Some taxa that failed the logical test, such as the shrubs *Eriogonum heermannii* and *E. microthecum* and the dicot herb *Stephanomeria spinosa,* had data that fit this pattern. However, the extent that woodrats select against these taxa is unknown and requires additional studies. Thus, for taxa that are avoided by woodrats, the probability estimates derived from the modern sample pairs (i.e., upper limit p(T) and lower limit p(F)) are reliable, which is similar to the conclusion reached in the preceding paragraph for taxa that occur much more frequently in modern pairs, but we do not expect that additional paleovegetation pairs would help resolve the data discrepancy. If woodrats selected *for* a taxon, this effect should have only a small influence on probability estimates because: (1) our method is based upon presence/absence data rather than abundance, and (2) the method already accounts for the state of false presence.

Although inferential and statistical methods are available to evaluate gaps in the fossil record (Blatt *et al.,* 1991), our method has four advantages compared to existing techniques. First, the evaluation of a particular taxon is made independent of features derived from characteristics of the stratum, such as from the paleoenvironment or from association with other taxa. The only requirement for our method is that the pair of samples is representative of the same locale and time period. Second, accuracy of the probability estimates generally can be increased simply by increasing sample size. As sample size increases, the upper and lower probability limits for p(F), as well as those for p(T), converge (except for taxa that are selected against by the woodrats). Third, the methodology provides an internal consistency check of the results, which also highlights taxa that may need additional sampling. Finally, the procedure has high reliability for both relatively abundant and relatively rare taxa.

Our method can be applied readily to other types of stratigraphic sampling, such as the analysis of macrofossils from sediment cores or of fossils from biostratigraphic units. Typically, broader application will require broader definitions of terms and analogous assumptions. In an example of a broader definition, where we would use the term "vegetation" others would use the term "landscape" to indicate the biota within a specific geographic area from which the fossils in a stratigraphic sample are derived. Thus, the landscape for midden samples is the area within 100 m of the midden, but the landscape encompasses the area within 30 m of the pond's shore for conifer macrofossils from sediment cores (Dunwiddie, 1987). An assumption analogous to "behavioral patterns of woodrats are the same as each midden in a paired sample is constructed" is "depositional and compressional characteristics are the same for both stratigraphic samples." Clearly, our technique is most useful for extant taxa. For extinct species, the method does not provide an internal check of the results, but it is still useful because the worst possible cases (lower limit of true inference and upper limit of false inference) are estimated.

ACKNOWLEDGMENTS

We thank J. Lyons-Weiler for help in formalizing the probability estimates; J. Allen, L. Bellis, C. Biggart, M. Brown, J. Menasco, T. Morgan, and M. Morrison for help with lab and field work; and R. Thompson and an anonymous reviewer for helpful comments on an earlier draft of the manuscript. This research was funded in part by a USDA NRICGP grant, the USDA Forest Service, and the Nevada Agricultural Experiment Station.

REFERENCES

Beck, M. W. (1996). On discerning the cause of the late Pleistocene megafaunal extinctions. *Paleobiology* **22,** 91–103.

Bennington, J. B., and Bambach, R. K. (1996). Statistical testing for paleocommunity recurrence: Are similar fossil assemblages ever the same? *Palaeogeography, Palaeoclimatology, Palaeoecology* **127,** 107–133.

Betancourt, J. L., Van Devender, T. R., and Rose, M. (1986). Comparison of plant macrofossils on woodrat (*Neotoma* sp.) and porcupine (*Erethizon dorsatum*) middens from the western United States. *Journal of Mammalogy* **67,** 266–273.

Blatt, H., Berry, W. B. N., and Brande, S. (1991). "Principles of Stratigraphic Analysis." Blackwell Scientific, Boston.

Boucot, A. J. (1996). Epilogue. *Palaeogeography, Palaeoclimatology, Palaeoecology* **127,** 339–359.

Brett, C. E., Ivany, L. C., and Schopf, K. M. (1996). Coordinated stasis: An overview. *Palaeogeography, Palaeoclimatology, Palaeoecology* **127,** 1–20.

Dial, K. P., and Czaplewski, N. J. (1990). Do woodrat middens accurately represent the animals' environments and diets? The Woodhouse Mesa study. *In* "Packrat Middens—The Last 40,000 Years of Biotic Change" (J. L. Betancourt, T. R. Van Devender, and P. S. Martin, Eds.), pp. 43–58. Univ. of Arizona Press, Tucson.

Dunwiddie, P. W. (1987). Macrofossil and pollen representation of coniferous trees in modern sediments from Washington. *Ecology* **68,** 1–11.

Foote, M., and Sepkoski, J. J., Jr. (1999). Absolute measures of the completeness of the fossil record. *Nature* **398,** 415–417.

Grayson, D. K., and Livingston, S. D. (1993). Missing mammals on Great Basin Mountains: Holocene extinctions and inadequate knowledge. *Conservation Biology* **7,** 527–532.

Holland, S. M. (1995). The stratigraphic distribution of fossils. *Paleobiology* **21,** 92–109.

Marshall, C. R. (1994). Confidence intervals on stratigraphic ranges: Partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology* **20,** 459–469.

McKinney, M. L., Lockwood, J. L., and Frederick, D. R. (1996). Does ecosystem and evolutionary stability exclude rare species? *Palaeogeography, Palaeoclimatology, Palaeoecology* **127,** 191–207.

Mensing, S. R., and Byrne, R. (1998). Pre-mission invasion of *Erodium cicutarium* in California. *Journal of Biogeography* **25,** 757–762.

Nowak, C. L., Nowak, R. S., Tausch, R. J., and Wigand, P. E. (1994a). A 30,000 year record of vegetation dynamics at a semi-arid locale in the Great Basin. *Journal of Vegetation Science* **5,** 579–590.

Nowak, C. L., Nowak, R. S., Tausch, R. J., and Wigand, P. E. (1994b). Tree and shrub dynamics in northwestern Great Basin woodland and shrub steppe during the Late-Pleistocene and Holocene. *American Journal of Botany* **81,** 265–277.

Popper, K. R. (1991). The non-existence of probabilistic inductive support. *In* "Logical Foundations: Essays in Honor of D. J. O'Conner" (I. Mahalingam and B. Carr, Eds.), pp. 31–41. St. Martin's Press, New York.

Solow, A. R. (1996). Tests and confidence intervals for a common upper endpoint in fossil taxa. *Paleobiology* **22,** 406–410.

Spaulding, W. G., Betancourt, J. L., Croft, L. K., and Cole, K. L. (1990). Packrat middens: Their composition and methods of analyses. *In* "Packrat Middens—The Last 40,000 Years of Biotic Change" (J. L. Betancourt, T. R. Van Devender, and P. S. Martin, Eds.), pp. 59–85. Univ. Arizona Press, Tucson.

Thompson, R. S. (1985). Palynology and *Neotoma* middens. *American Association of Stratigraphic Palynologists Contribution Series* **16,** 89–112.

Thompson, R. S. (1990). Late-Quaternary vegetation and climate in the Great Basin. *In* "Packrat Middens—The Last 40,000 Years of Biotic Change" (J. L. Betancourt, T. R. Van Devender, and P. S. Martin, Eds.), pp. 200–239. Univ. of Arizona Press, Tucson.

Wells, P. V. (1983). Paleobiogeography of montane islands in the Great Basin since the last glaciopluvial. *Ecological Monographs* **53,** 341–382.