INFERRING MICROHABITAT PREFERENCES OF Lilium catesbaei (Liliaceae)\textsuperscript{1}

KRISTEN PENNEY SOMMERS\textsuperscript{2,4}, MICHAEL ELSWICK\textsuperscript{2,3}, GABRIEL I. HERRICK\textsuperscript{3}, and GORDON A. FOX\textsuperscript{3,5}

\textsuperscript{2}Pinellas County Biological Field Station, Pinellas County Department of Environmental Management, Tarpon Springs, Florida 34688, USA; and \textsuperscript{3}Department of Integrative Biology, University of South Florida, Tampa, Florida 33620-5150, USA

\footnotesize{\textsuperscript{1}Present address: Florida Fish and Wildlife Conservation Commission, Division of Habitat and Species Conservation, 620 S. Meridian Street, Tallahassee, Florida 32399-1600, USA.}

\footnotesize{\textsuperscript{2}Department of Forestry and Natural Resources, University of Florida, Gainesville, Florida 32611, USA.}

\footnotesize{\textsuperscript{3}Department of Integrative Biology, University of South Florida, Tampa, Florida 33620-5150, USA.}

\footnotesize{\textsuperscript{4}Present address: Florida Fish and Wildlife Conservation Commission, Division of Habitat and Species Conservation, 620 S. Meridian Street, Tallahassee, Florida 32399-1600, USA.}

\footnotesize{\textsuperscript{5}Author for correspondence (e-mail: gfox@usf.edu)}

\begin{itemize}
  \item \textit{Premise of the study:} Microhabitat studies use varied statistical methods, some treating site occupancy as a dependent and others as an independent variable. Using the rare \textit{Lilium catesbaei} as an example, we show why approaches to testing hypotheses of differences between occupied and unoccupied sites can lead to erroneous conclusions about habitat preferences. Predictive approaches like logistic regression can better lead to understanding of habitat requirements.
  \item \textit{Methods:} Using 32 lily locations and 30 random locations >2 m from a lily (complete data: 31 lily and 28 random spots), we measured physical conditions—photosynthetically active radiation (PAR), canopy cover, litter depth, distance to and height of nearest shrub, and soil moisture—and number and identity of neighboring plants. Twelve lilies were used to estimate a photosynthetic assimilation curve. Analyses used logistic regression, discriminant function analysis (DFA), (multivariate) analysis of variance, and resampled Wilcoxon tests.
  \item \textit{Key results:} Logistic regression and DFA found identical predictors of presence (PAR, canopy cover, distance to shrub, litter), but hypothesis tests pointed to a different set (PAR, litter, canopy cover, height of nearest shrub). Lilies are mainly in high-PAR spots, often close to light saturation. By contrast, PAR in random spots was often near the lily light compensation point. Lilies were near \textit{Serenoa repens} less than at random; otherwise, neighbor identity had no significant effect.
  \item \textit{Conclusions:} Predictive methods are more useful in this context than the hypothesis tests. Light availability plays a big role in lily presence, which may help to explain increases in flowering and emergence after fire and roller-chopping.
\end{itemize}

\textbf{Key words:} discriminant function analysis; \textit{Lilium catesbaei}; logistic regression; microhabitat preferences; photosynthetically active radiation (PAR); safe site; spatial dispersion.

The patterns and causes of dispersion of plant populations in space and time are fundamental issues in population biology because dispersion affects population and community dynamics and, consequently, is important in management efforts, in applications including conservation, pest management, and forestry and range science. There has been much recent interest in the mechanisms that affect the dispersion of plant populations (see review in Münzbergová and Herben, 2005). Generally, spatial pattern is caused by differences among the spatial scales of habitat availability, dispersal, competition, and natural enemies. For example, clumping on small spatial scales may be a consequence of limited seed dispersal. Individuals become clumped because seedfall is near the maternal plant (Eriksson and Ehrlén, 1992; Losos, 1995; Hilton and Boyd, 1996; Eriksson, 1998; Wolf et al., 1999; Ehrlén and Eriksson, 2000; Kupferschmid et al., 2000; Turnbull et al., 2000; Coulson et al., 2001; Gustafsson et al., 2002; Jacquemyn et al., 2002; Tofts and Silvertown, 2002), while other suitable habitat remains unoccupied (Jones et al., 2006; Rey et al., 2006). Clumped dispersion of plants in space may also result if seed predation (Orrock et al., 2006), herbivory (e.g., Boyd, 1988; Ehrlén, 1996; Gomez, 2005), or disease transmission (Burdon and Chilvers, 1982) occurs mainly on small spatial scales in relation to dispersal. Finally, suitable habitat for seed deposition, establishment, or subsequent survival and growth may itself be patchy, resulting in microsite limitation. None of these factors are mutually exclusive (Moore and Elmdendorf, 2006).

Understanding the processes that create the spatial distribution of a species may be even more important in rare organisms that may not occupy all available suitable habitat (Hilton and Boyd, 1996) or that may be more susceptible to local extirpation due to environmental and demographic stochasticity than common widespread species. Small herbaceous plants may be especially sensitive to microsite variation, particularly in habitats with dominant shrub and tree species (Menges and Kimmich, 1996). These larger species restrict the availability of light, water, soil resources, and physical space. Decisions on best management techniques to support or increase population sizes of rare plants depend on understanding the contribution of seed dispersal versus microsite limitation in creating an observed pattern in nature; however, ecological studies of rare species rarely provide enough information to understand the demographic or ecological consequences of those management techniques (Doak and Mills, 1994, Kephart and Paladin, 1997; however see Menges and Kimmich, 1996; Thompson et al., 2006). Here, we examine the role of microsite variation in explaining the observed pattern of a rare perennial of the southeastern United States as part of a long-term demographic study in managed pine flatwoods.

Many empirical plant microhabitat studies have focused on the relationship between microsite variation and seedling establishment. Sites suitable for germination and establishment...
are called “safe sites,” and the degree of suitability of habitat at a small scale may vary across a site or within available habitat (sensu Harper et al., 1961; Fowler, 1988; Allison and Ehrenfeld, 1999). Establishment and survival to maturity may be a result of variation in light (Castellanos et al., 1999), litter and soil fauna (Kalliovirta et al., 2006; Negrete Yankelevich et al., 2006), physical soil factors such as soil moisture, depth or temperature (Franco and Nobel, 1989; Colling et al., 2002), disturbance (Menges and Hawkes, 1998), or a combination of factors. Several studies have focused on the evolution of the characteristics of safe sites for different species (Jaenike and Holt, 1991; Martin, 1998).

Microsite variation also affects mature plants. Physical characteristics of microsites may affect reproductive success (Agren, 1988; Escarre and Thompson, 1991; Menges and Kimmich, 1996). In perennial plants, preferences in microhabitat may also affect long-term adult survivorship (Kephart and Paladino, 1997). Perennial plants may be able to mitigate short-term stochastic changes in microsite variation with multiple flowering seasons that ensure higher probabilities of reproductive success in some years. Perennial species also may evolve mechanisms for temporal habitat choice through adult dormancy periods; therefore, because of their potential long-lived nature, perennial plants may be able to escape short-term stochastic changes in microsite suitability by reserving resources or becoming dormant while conditions are unfavorable (Bazzaz, 1991; Lesica and Steele, 1994; Shefferson et al., 2003; Kéry and Gregg, 2004; Lesica and Crane 2007). Our focus here is on adult plants.

Statistical issues—What are the best strategies for identifying the factors that affect microhabitat use? A natural approach—but one that we contend can easily be misleading—is based on classical hypothesis-testing. In a nutshell, the idea is to ask (with observations or experiments) whether “good” locations (determined by presence of the study species) differ from “poor” locations (determined by absence). For example, a researcher or manager might test whether mean soil pH differs between occupied and unoccupied sites. However, there is a conceptual problem: testing the hypothesis that occupied and unoccupied locations are different (in the mean or rank of some characteristic) is not equivalent to asking whether that characteristic is a useful predictor of the presence or absence of the plant species.

There are several reasons why these approaches are not equivalent. The first reason is mathematical and is easily seen by comparing an analysis of variance (ANOVA) testing for differences between soil pH in occupied versus unoccupied sites with a logistic regression using soil pH to predict occupancy. The ANOVA models the mean pH as a function of occupancy, while the logistic regression models the odds of occupancy as a function of soil pH. More precisely, the ANOVA models $E(y|x)$, where $E$ is for expectation, $y$ is soil pH (in this case), and $x$ is the treatment variable, occupancy. By contrast, the logistic regression models the log-odds

$$
\frac{p}{1-p}
$$

where $p$ is the probability of occurrence. Thus, the conditioning of the models is different: the ANOVA estimates mean soil pH conditioned on occupancy, whereas the logistic regression estimates the probability of occupancy, conditioned on observed soil pH.

Their orthogonal conditioning means that statistical inference is rather different for these two models. In the ANOVA model, inferences concern differences between the mean soil pH values, given the within- and between-treatment variances. In the logistic regression, inferences concern whether prediction of occupancy is improved by information on soil pH. The differences between two mean values may not be greater than expected at random (by ANOVA) even though soil pH is a useful predictor of occupancy, and the reverse can also be true.

These models also differ considerably in their underlying assumptions. An ANOVA requires that residuals be normally distributed, with equal variances. The closest nonparametric analogue—the Mann-Whitney $U$-test—makes a somewhat weaker assumption, that the two distributions have the same (arbitrary) shape. No such assumptions are necessary for logistic regression (Floyd, 2001). Thus, logistic regression models are able to make use of the entire distribution of observed data, rather than simply the means. It is easy to imagine situations in which two treatments have the same mean but different variances; perhaps one is also skewed. An ANOVA or Mann-Whitney test may lead one to the conclusion that the distributions are not different, but it may be that the other differences between the distributions allow logistic regression to predict occupancy.

The final reason why these approaches are not equivalent is that the problem is typically multivariate. Imagine that we now want to study the relationship between site occupancy on the one hand, and soil pH and light availability on the other. We can conduct a multivariate analysis of variance (MANOVA), but if our analysis says that there are significant treatment effects, how do we decide whether the differences involve soil pH, light availability, or both? Many authors (e.g., Scheiner, 2001) suggest conducting a canonical analysis of the MANOVA for this purpose. This canonical analysis is simply another name for discriminant function analysis (DFA), which is designed to find the combination of variables that best distinguishes between groups. In other words, as soon as we want to make inferences about more than a single factor, we need a method designed for inferences about how to distinguish between occupied and vacant sites. An alternative approach widely employed in ecology is to use a series of ANOVAs to make inferences about which variables contribute to the differences between sites. But this alternative takes us back to all the problems discussed above. In general, then, we suggest that the (M) ANOVA approach will often not do the job we want. We recommend using methods designed to predict presence versus absence (or to cluster or discriminate between locations), such as logistic regression and DFA.

Both general approaches have been used in microhabitat studies (Fowler, 1988; Hawkes and Menges, 1995; Menges and Kimmich, 1996; Kephart and Paladino, 1997; Scheiner, 2001; Ibáñez and Schupp, 2002), sometimes even in the same study. However, the fundamental difference between these approaches has not been widely appreciated. Here, we explicate the differences between the methods by using both approaches in studying the microhabitat requirements of a lily.

Study organism—Catesby’s lily (Lilium catesbaei Walter) is a poorly studied species listed as threatened by the state of Florida (Coile and Garland, 2003). This species has the largest flower of any North American lily (Skinner, 2002). In our study area, flowering occurs in late August through November, and
seeds are dispersed from October to January. A mature *L. catesbaei* plant emerges from a 1–3 cm underground bulb and consists of a stem that may reach over a half meter, with cauline leaves, long basal leaves, and one large red-orange flower (Huffman, 1997). A three-part capsule develops that may contain >500 small paper-thin seeds, typically ~0.5 cm in length. Germination occurs during the winter months. Seedlings may take up to 2 years before producing a reproductive stem. As in many long-lived perennial herbs (Hutchings, 1987; Alexander et al., 1997), individuals sometimes remain below ground during some years. In years when aboveground parts are present, these can be only basal leaves, basal leaves and a non-bud-producing stem, or a bud-producing stem. *Lilium catesbaei* are known to emerge in large numbers and flower after fire events (Huffman, 1997; Huffman and Werner, 2000). With fire suppression, they may disappear or lie dormant (Huffman and Werner, 2000), similar to other plants in fire-maintained ecosystems (Brewer and Platt, 1994). Little is known about the general ecology, population biology, or microhabitat requirements of *L. catesbaei* (Abrahamson and Hartnett, 1990; Huffman and Werner, 2000).

The range of *L. catesbaei* includes pine flatwoods and other open, moist habitats throughout southeastern North America, but it is locally rare, possibly because of fire suppression or habitat degradation. Huffman and Werner (2000) compared its responses to prescribed fire, restoration roller-chopping (a vegetation management method in which a weighted drum with blades on the outside is towed through an area), and a combination of roller-chopping followed by prescribed fire, and found many more flowering plants after roller-chopping than after the other treatments. Their results further suggested that the number of flowering plants is negatively correlated with the amount of cover by saw palmetto (*Serenoa repens*). However, they did not attempt further explanation of this response; it is unclear whether it is due to changes in nutrient availability, light levels, litter, or other factors. Understanding the contribution of microhabitat factors to the observed spatial distribution of this species may be crucial to understanding its rare and declining occurrence in pine flatwoods and to determining what management techniques may help or hinder its presence across its range. In the present study, we aimed to understand the factors associated with presence of flowering lilies, both as an aid to management and to gain insight into how populations persist in the face of spatial and temporal heterogeneity. The observed spatial clustering and increased emergence and flowering after fires suggested the possibility that changes in light availability might be a key for this species, and so we pay special attention to factors related to light availability.

**METHODS**

**Study site**—In 1995 the Pinellas County (Florida) Environmental Lands Division began an effort to monitor the population of *L. catesbaei* at the Brooker Creek Preserve (BCP), a 3400-ha preserve near Tarpon Springs, Florida. Over the ensuing years, >1500 individuals were identified during exhaustive searches in five natural flatwoods site and several specially managed sites. They were marked and mapped with high-precision GPS equipment and recensused annually (K. P. Sommers, unpublished data). One striking pattern in the BCP data set is that flower production varies greatly among years and sites (Fig. 1). With *L. catesbaei* individuals clustered spatially (Fig. 2) within these sites. For the present study, we selected one of these sites, in ~4.98 ha of open pine flatwoods, bordered by a cypress swamp and heavily disturbed powerline easements. Vegetation was typical of a pine flatwoods, with a sparse canopy of slash pine (*Pinus elliottii var. densa*) and a dense shrub layer consisting of *S. repens*, gallberry (*Ilex glabra*), staggerbush (*Lyonia fruticosa*), dwarf live oak (*Quercus minima*), wax myrtle (*Myrica carterae*) and creeping huckleberry (*Gaylussacia nana*); nomenclature is according to Wunderlin and Hansen (2003).

**Sampling design**—In May 2003, thirty-two individual lilies that flowered in the previous fall were selected at random from our database of marked plants. For comparison, 30 randomly located vacant points were also selected. Vacant points were located using the ArcView 3.3 (ESRI, Redlands, California, USA) extension “Random Point Generator v1.27” such that they were selected within a polygon formed by the external boundaries of the study site. Vacant points in all cases were located ≥2 m from the nearest known lily and ≥2 m from any other selected point. Vacant points were located in the field using a mapping-grade GPS unit (Trimble ProXR; Trimble Navigation, Sunnyvale, California, USA).

Microhabitat variables were measured for each plant or point within the cylindrical space formed within a 1-m circular plot. This area corresponds to the maximum horizontal space that lily individuals with a flowering stem have been observed to occupy. Microhabitat parameters included both biotic and abiotic factors. Photosynthetically active radiation (PAR) was estimated using a Li-Cor LI-190SA Quantum Sensor (Li-Cor Biosciences, Lincoln, Nebraska, USA) placed 3 cm above the surface, a distance that corresponds to the average midleaf height of basal leaves. Two measurements were taken per plant, on the same day. Data were taken during a 2-wk period in May on days with little or no cloud cover between 1100 and 1300 hours. Percent canopy cover was measured above the plot center at a height of 3 cm using a spherical densiometer. Measurements were taken in the four cardinal directions, and the mean of the four spherical densiometer readings was used to estimate percent canopy cover for each plot. Soil moisture was assessed by collecting soil cores 10–20 cm deep, located 15–50 cm from the plot center, using an Oakfield model LS soil sampler (Oakfield Apparatus, Oakfield, Wisconsin, USA). Care was taken so that sampling would not damage lilies but still be representative of soil conditions within their rooting zone. All soil samples were collected within the same morning period within one day. The sampler was used by removing any litter from the surface of the soil, plugging the probe into the ground and immediately placing the soil sample into a plastic bag. The bag was then sealed for transportation to the laboratory. Soil samples were weighed, dried using a standard laboratory oven, and reweighed to gravimetrically determine water content. Litter depth (cm) was also measured by selecting four representative points within plots. These points were selected so that they were spaced throughout the plot and did not disturb the lily (near lily points) or nearby shrubs. The mean of these four values was used in analyses.
Maximum shrub height within each quadrant and the distance to and identity of the nearest shrub were recorded. Because *S. repens* was the most dominant shrub throughout the site, we estimated specifically its percent cover within each plot.

To examine the photosynthetic response to light intensity, we selected 12 mature *L. catesbaei* in our study site. We selected plants that were large enough to measure, had aboveground vegetative parts during this sampling, and were accessible to our equipment. Using a Li-Cor 6400 IR gas analyzer (Li-Cor Biosciences), we measured carbon dioxide assimilation rate at 10 photosynthetic photon flux densities from 1400 down to 0 μmol s⁻¹ using the “LightCurve” autoprogram. Light level was manipulated using the built-in lamp in the portable gas chamber. Leaves were exposed to new light levels for a minimum of 120 s and a maximum of 200 s to allow photosynthesis to stabilize. Temperature, humidity, and CO₂ concentration were held constant throughout sampling.

**Data analysis**—Complete data were available for 31 lily and 28 random locations. To normalize errors, we log-transformed data on soil moisture, litter depth, PAR, and shrub height and square-root transformed data on distance to nearest shrub. Where it was desirable to interpret parameter estimates as the relative importance of particular variables, we further standardized the data by subtracting the means and dividing by the standard deviation of each variable.

We used multivariate and univariate ANOVAs to ask whether lily-occupied and randomly chosen spots differed in physical factors. We conducted a discriminant function analysis (DFA) to examine which variables contributed heavily to the MANOVA results. To ask whether our results flowed from the particular assumptions of ANOVA, we also conducted Mann-Whitney *U*-tests for these hypotheses. Because of multiple ties, we used a permutation approach (Hothorn et al., 2008) to estimate *P* values.

We examined the correlation structure of these data with both Pearson product-moment correlations and a principal component analysis.

To examine the importance of these physical factors, we used multiple logistic regression to find the best model for predicting lily presence. After estimating a model including all physical predictors, we used Akaike’s Information Criterion (AIC) to choose between models using different subsets of the predictors. The logistic regression and DFA approaches ask similar questions; however, logistic regression requires many fewer assumptions. Discriminant function analysis makes the same assumptions as MANOVA—especially multivariate normality and homogeneity of variances; neither assumption is needed for logistic regression. On the other hand, when these assumptions are met, DFA has greater power. Logistic regression can accommodate categorical values—which DFA cannot—and the coefficients estimated by logistic regression have a somewhat more natural interpretation.

For data collected using cover classes, chi-square tests were used to compare lily-occupied with randomly selected spots, including herbaceous cover, nearest shrub species, and *S. repens* cover.

We estimated the photosynthetic response curve by using nonlinear regression to fit (following Peek et al., 2002) the equation

$$A = A_{\text{max}} \left[ 1 - e^{-A_q \left( \text{PPF} - \text{LCP} \right)} \right]$$

where *A* is the assimilation rate, *A*ₘₐₓ is the estimated maximum assimilation rate, *A*ₚ is the estimated apparent quantum yield, PPF is the photosynthetic photon flux, and LCP is the estimated light compensation point.

We used Systat version 9.0 (SPSS, Chicago, Illinois, USA) for the vegetation analyses, and R software (R Project Core Development Team, Vienna, Austria) version 2.11 for all other analyses.

**RESULTS**

**Physical characteristics**—Lily-occupied spots differed from randomly selected spots in most variables measured (Fig. 3). Lilies tended to occupy more open sites, as indicated by the smaller means and medians for canopy cover, greater PAR, and smaller maximum shrub height, as shown by both ANOVA and nonparametric tests (Fig. 3). This trend is also suggested by the lower litter depth for lily sites. None of the univariate tests found a difference in distance to nearest shrubs between lily-occupied and randomly selected spots, though Figure 3 clearly
shows that there are more randomly selected than lily-occupied spots with large distances to the nearest shrub. Lilies tended to occupy areas of a smaller range of available moisture within our study site, with a lower mean; however, the differences in mean and rank were not statistically significant. Percent bare ground was also analyzed, but because these data could not be transformed to meet normality assumptions, we used a Mann-Whitney test, which provided a weak suggestion of a difference in the amount of bare ground in lily occupied versus randomly selected spots ($U = 608$, df = 1, $P = 0.08$).

The correlations among physical variables are shown in Figure 4 (top). In general, these correlations are small in magnitude; the largest in magnitude are those between PAR and litter depth ($r = -0.38$) and between canopy cover and shrub height ($r = 0.39$). Interestingly, the correlation between PAR and canopy cover was somewhat smaller, with $r = -0.24$, which may suggest the importance of indirect sources of light in this habitat.

Two principal components were significant; their loadings are given in Figure 4 (middle). The first component (accounting for 34% of the variance) involves all terms, with PAR and distance to shrub having negative loadings whereas all others are positive. The moisture and distance to shrub loadings were somewhat smaller in magnitude than the others. The second principal component (20% of the variance) was heavily loaded by distance to shrub, along with litter and moisture, all of which loaded negatively. Scores on the first two principal components are shown in Figure 4 (bottom). Lily-occupied spots mainly
occurred around low values of PC1 and high values of PC2, whereas randomly selected spots occurred more widely throughout the principal component space.

The full logistic regression (Table 1) suggests that PAR, canopy cover, and distance to shrub were the most important factors in predicting lily presence. A model with only four predictors (PAR, litter depth, percent canopy cover, and distance to shrub) was the best-supported model (Table 1). PAR and canopy cover had the strongest effects, whereas the effects of distance to shrub and litter depth were roughly as strong as one another. The DFA gave results very similar to those of the logistic regression, indicating PAR as the most important single factor, distance to shrub and canopy cover as having smaller but similar importance, followed by litter depth, shrub height, and moisture, respectively. Figure 5 shows that both the best-fitting logistic regression and the discriminant function succeeded well in separating lily-occupied from randomly selected points.

Results of the photosynthetic analysis are shown in Figure 6. The light compensation point is estimated to be at a photosynthetic photon flux density of 34.12 μmol photons m⁻² s⁻¹. By comparison, the mean available PAR in lily occupied sites, the mean net fixation rate of 10.49 μmol CO₂ cm⁻² s⁻¹. Using the mean available PAR in lily occupied sites, the mean net fixation rate of 10.49 μmol CO₂ cm⁻² s⁻¹ is close to the asymptote, 11.28 μmol CO₂ cm⁻² s⁻¹. By comparison, the mean available PAR in randomly selected sites gives a mean fixation rate of 4.08 μmol CO₂ cm⁻² s⁻¹, which is fairly close to the compensation point.

Vegetation analysis—Twenty-eight species of herbaceous plants were observed as being dominant within plots at our study site (Table 2). Because of the low frequency of herbaceous plants recorded in each plot, we examined the four most frequently occurring herbaceous plant species. *Andropogon glomeratus*, *Aristida purpurascens* var. *virgata*, *Ar. stricta*, and *Panicum* spp. were analyzed using separate chi-square tests with cover category as the frequency category. None of these species occurred in different proportions of cover class frequencies between the lily-occupied and randomly selected spots (A. *glomeratus*: χ² = 5.433, df = 5, P > 0.1, n = 17; *Aristida purpurascens* var. *virgata*: χ² = 7.584, df = 5, P > 0.1, n = 38; *Ar. stricta*: χ² = 1.428, df = 5, P > 0.1, n = 29; *Panicum* spp.: χ² = 7.189, df = 5, P > 0.1, n = 29.). Identity of nearest shrub species recorded also had low frequency. Because of this low frequency, Pearson chi-square analyses (with corrections for small expected values) were completed only for those plants recorded more than four times (Table 2). These species included *I. glabra*, *L. fruticosa*, *M. cerifera*, *S. repens*, and *Vaccinium arboetum*. Analyses indicated that the relative frequency of shrub species closest to where lilies occurred (Table 3) was not different from that at locations not occupied by lilies (χ² = 4.584, df = 4, P = 0.333, n = 57). *Serenoa repens* was the most dominant species within the shrub layer, and the amount of *S. repens* cover was categorized for each study plot. *Serenoa repens* cover was classified more often in the higher-percentage cover categories in the randomly selected spots than in the lily-occupied plots (χ² = 15.202, df = 5, P < 0.01, n = 62), which indicates that *S. repens* cover was lower in areas that included lilies.

### DISCUSSION

There is no single best approach to the statistical problem of how to find the factors that are important for plant microhabitat use. Different approaches will be more or less useful, depending on the biological and statistical details of the problem. While it seems natural to ecologists to reword the question “what predicts site occupancy” as “do occupied and vacant sites differ in mean or rank order of this feature,” our results point to the conclusion that the latter question can often lead to mistaken inferences.

Using either ANOVA or resampled Mann-Whitney U-tests, we identified differences between occupied and unoccupied sites in PAR, litter depth, canopy cover, and shrub height, whereas logistic regression and DFA identified PAR, litter depth, distance to shrub, and percent canopy cover as the most important microhabitat features. There are two major reasons why these approaches identified different sets of characteristics as important. First, distance to shrub is a key part of the correlation structure, as the principal component analysis makes clear (Fig. 4). Logistic regression and DFA allow us to examine all the site characteristics simultaneously, which is not so for either the ANOVA or Mann-Whitney approaches. We can test for significance simultaneously using MANOVA, but then ask which characteristics contribute to the differences requires using the DFA. The second reason for the difference in statistical results is that the shapes of the distributions differ for distance to shrub and shrub height (Fig. 3). There is little difference between the mean distance to shrub of occupied versus unoccupied sites, but the unoccupied sites have a long tail, with little overlap from the occupied sites. Distance to shrub becomes a useful predictor of occupancy as a result. On the other hand, the proportionally greater overlap between the central quartiles of the distributions for shrub height in occupied and unoccupied sites (Fig. 3) makes this variable less useful in distinguishing the site types, despite the proportionally greater difference between their means.

Although we succeeded in identifying a set of physical variables to predict occupancy, our approach could have failed had any of several conditions applied. If suitable sites were common but lilies occupied only a small fraction of them—because of dispersal limitation or some other factor limiting population

---

**Table 1.** Logistic regression and discriminant function analysis of physical variables. LR is the likelihood ratio and P its significance in a Type II analysis of deviance. All variables transformed as in text.

<table>
<thead>
<tr>
<th>Physical variable</th>
<th>Standardized discriminant function coefficient</th>
<th>Coefficient (SE)</th>
<th>LR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil moisture (%)</td>
<td>0.0884</td>
<td>-0.2618 (0.5014)</td>
<td>0.268</td>
<td>0.605</td>
</tr>
<tr>
<td>Shrub height (cm)</td>
<td>0.1775</td>
<td>-0.4006 (0.4404)</td>
<td>0.840</td>
<td>0.359</td>
</tr>
<tr>
<td>PAR (μmol m⁻² s⁻¹)</td>
<td>-0.7722</td>
<td>1.7504 (0.5501)</td>
<td>17.254</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Litter depth (cm)</td>
<td>0.2098</td>
<td>-0.8717 (0.5825)</td>
<td>2.644</td>
<td>0.104</td>
</tr>
<tr>
<td>Canopy cover (%)</td>
<td>0.4178</td>
<td>-1.3887 (0.6942)</td>
<td>4.602</td>
<td>0.032</td>
</tr>
<tr>
<td>Distance to shrub (cm)</td>
<td>0.5240</td>
<td>-1.0475 (0.5164)</td>
<td>4.982</td>
<td>0.026</td>
</tr>
</tbody>
</table>

---

**Table 2.** Characteristics contributing to the difference in utilization of lily occupied and unoccupied sites.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>LR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil moisture</td>
<td>0.268</td>
<td>0.605</td>
</tr>
<tr>
<td>Shrub height</td>
<td>0.840</td>
<td>0.359</td>
</tr>
<tr>
<td>PAR</td>
<td>17.254</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Litter depth</td>
<td>2.644</td>
<td>0.104</td>
</tr>
<tr>
<td>Canopy cover</td>
<td>4.602</td>
<td>0.032</td>
</tr>
<tr>
<td>Distance to shrub</td>
<td>4.982</td>
<td>0.026</td>
</tr>
</tbody>
</table>

---

**Table 3.** Characteristics determining the difference in proportion of cover categories in lily occupied and unoccupied plots.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>LR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil moisture</td>
<td>0.268</td>
<td>0.605</td>
</tr>
<tr>
<td>Shrub height</td>
<td>0.840</td>
<td>0.359</td>
</tr>
<tr>
<td>PAR</td>
<td>17.254</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Litter depth</td>
<td>2.644</td>
<td>0.104</td>
</tr>
<tr>
<td>Canopy cover</td>
<td>4.602</td>
<td>0.032</td>
</tr>
<tr>
<td>Distance to shrub</td>
<td>4.982</td>
<td>0.026</td>
</tr>
</tbody>
</table>
Fig. 5. Goodness-of-fit for the logistic regression and discriminant function analysis. (A) Ln[abs(residuals)] from best-supported logistic regression; dashed line is the mean of fitted values. (B) Boxplots for discriminant function scores.

Fig. 6. Net assimilation rate as a function of light intensity in Lilium catesbaei. The mean photosynthetic photon flux density (PPFD) in lily-occupied sites is near the estimated light saturation point for this population, whereas the mean PPFD in randomly selected sites is near the estimated light compensation point.

on mature plants—and not on “safe sites”—the evidence, taken together, suggests that germination and establishment may require similar conditions. The logistic regression and DFA resulted in the same sets of predictors but weighted them in slightly different order. This difference may not have practical significance, but we note that logistic regression is a less assumption-laden approach than DFA for this kind of problem.

Our results strongly suggest that L. catesbaei prefer open, high-light locations with minimal litter. Indeed, these results suggest that the availability of such sites may play a key role in explaining why these plants are so strongly clustered in space (Fig. 2), why so many of them appear and flower in the year or so following fires, and why they have so strong a response to roller-chopping (Huffman and Werner, 2000). Extensive studies have examined the relationship between open gaps in the canopy and subcanopy in the tropics caused by natural treefall (Brokaw, 1985; Dalling et al., 1998), and in Florida scrub habitat caused by fire (Hawkes and Menges, 1995; Menges and Kimmich, 1996; Petru and Menges, 2003). Many other studies have specifically examined the role of fire, the resulting increase in light levels, and increased flowering (Keeley et al., 1981; Ruiters et al., 1993; Tyler and Borchert, 2003; Himan and Brewer, 2007). Our results are consistent with other herbaceous species that depend on openings in the canopy produced by fire in pyrogenic systems.

Lilies tended to occupy areas with less S. repens cover than would be expected at random. Our other results on vegetative correlates do not point to differences between lily-occupied and randomly selected spots in the proportion of those plots that contained one of the four dominant herbaceous species. Nor did we find differences in the identity of nearby shrubs (beyond S. repens) between lily-occupied and randomly selected spots. It is possible that a larger data set might reveal differences too small for our analyses to detect.
Leaf litter has been found to help shape many other plants’ preferred microhabitat (e.g., Ibáñez and Schupp, 2002). Litter cover may have positive or negative consequences for germination (Facelli and Pickett, 1991a; McConnell and Menges, 2002; Quested and Eriksson, 2006; Myster, 2006), survival (Fowler, 1988), or flowering (Kalliovirta et al., 2006), and this may vary between species or over time (Kalliovirta et al., 2006; Myster, 2006). For L. catesbaei, litter may have a direct effect on the ability of basal leaves to emerge, or it may alter soil characteristics. Conversely, litter may simply be a consequence of the amount of vegetative cover in the area by affecting the light transmittance to the soil (Facelli and Pickett, 1991b).

**Conservation implications**—Physical structure in a natural pine flatwoods is driven primarily by fire. Fire reduces the amount of ground litter and undergrowth. In fire-free periods, the shrub overstory becomes dominated by woody species such as oaks (Quercus spp.) and other low shrubs and S. repens

<table>
<thead>
<tr>
<th>Table 2. Frequencies of dominant herbaceous plants within study plots.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lily-occupied plots</strong></td>
</tr>
<tr>
<td>Andropogon glomeratus</td>
</tr>
<tr>
<td>Aristida purpurascens var. virgata</td>
</tr>
<tr>
<td>Aristida stricta</td>
</tr>
<tr>
<td>Asimina reticulata</td>
</tr>
<tr>
<td>Carpephorus spp.</td>
</tr>
<tr>
<td>Centella asiatica</td>
</tr>
<tr>
<td>Eleocharis vivipara</td>
</tr>
<tr>
<td>Eleocharis elata</td>
</tr>
<tr>
<td>Eriogonum vernum</td>
</tr>
<tr>
<td>Eupatorium mohrii</td>
</tr>
<tr>
<td>Eupatorium rotundifolium</td>
</tr>
<tr>
<td>Euthamia caroliniana</td>
</tr>
<tr>
<td>Euthamia minor</td>
</tr>
<tr>
<td>Lachnanthes caroliniana</td>
</tr>
<tr>
<td>Lilium catesbaei</td>
</tr>
<tr>
<td>Panicum spp.</td>
</tr>
<tr>
<td>Paspalum spp.</td>
</tr>
<tr>
<td>Pilolepis rigida</td>
</tr>
<tr>
<td>Petrosedon esculentum</td>
</tr>
<tr>
<td>Rheza nutallii</td>
</tr>
<tr>
<td>Rynchospora spp.</td>
</tr>
<tr>
<td>Solidago fistulosa</td>
</tr>
<tr>
<td>Stillania sylvestra</td>
</tr>
<tr>
<td>Toxicodendron radicans</td>
</tr>
<tr>
<td>Vitis rotundifolia</td>
</tr>
<tr>
<td>Woodwardia virginica</td>
</tr>
<tr>
<td>Xyris brevifolia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Frequencies of shrubs closest to center point of study plots.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species</strong></td>
</tr>
<tr>
<td>Ilex glabra</td>
</tr>
<tr>
<td>Lyonia fruticosa</td>
</tr>
<tr>
<td>Myrica cerifera</td>
</tr>
<tr>
<td>Quercus spp.</td>
</tr>
<tr>
<td>Serenoa repens</td>
</tr>
<tr>
<td>Vaccinium arboreum</td>
</tr>
<tr>
<td>Vaccinium myrtillus</td>
</tr>
<tr>
<td>Hypericum tetrapetalum</td>
</tr>
<tr>
<td>Gaylussacia nana</td>
</tr>
</tbody>
</table>

(Litzenstein et al., 1995). Likewise, herbaceous cover and diversity may decline under the higher densities of hardwood and shrubs (Lemon, 1949; Abrahamson and Hartnett, 1990; Platt, 1999; Drewa et al., 2002). The reduction of shrub canopy—particularly S. repens—after fire may be the primary cause of gap openings, and thus of higher levels of light penetration into previously overgrown pine flatwoods. All three factors negatively associated with Catesby’s lilies are related to the amount of physical vegetative structure within its habitat. Adult dormancy exhibited by L. catesbaei may have evolved in this unpredictable habitat as a way to survive fire-free periods (reviewed in Bazzaz, 1991). Catesby’s lilies exhibit this type of temporal habitat choice, whereby individuals may become dormant for as much as 5 yr (K. P. Sommers, unpublished data). As the length of time since last emergence grows longer, the probability of survival by dormant individuals may decrease (e.g., Cochran and Ellner, 1992; Shefferson et al., 2001; Kéry and Gregg, 2004, Knight, 2004).

As pine flatwoods habitat becomes less available and current preserved pinelands are restored, basic understanding of the effects of management techniques will be crucial in ensuring that populations of rare and protected species are viable and protected in perpetuity. Restoration efforts or recovery of threatened plants via potential reintroduction attempts are dependent on being able to determine what constitutes suitable habitat.

Roller-chopping and fire have been shown to induce a flowering response in this species (Huffman and Werner, 2000). Our results suggest that this increase in flowering after these types of management treatments may be primarily a response to light. Management techniques that eliminate more litter and increase light availability (particularly prescribed fire and possibly roller-chopping) may provide more habitat for L. catesbaei. Because the plants are known to bloom in large numbers after fire and are cryptic when not in flower, prescribed fires may well reveal larger population sizes than previously known. Likewise, other management techniques that specifically reduce the amount of S. repens cover, such as herbicide or other mechanical removal, may have the most benefit to this species in areas where it is or should be found.

Sites more suitable for lilies at one time may tend to become less so as the canopy closes and litter depth increases. Improving our demographic understanding will require understanding the temporal and spatial scales over which these microhabitat factors vary, and the specific effects that they have on vital rates. Understanding the long-term relationship between habitat variability and demographic success may be crucial in long-term survival of this and other rare plants. This points to a final reason to prefer models predicting occupancy over hypothesis tests about differences between occupied and unoccupied sites: predictive models may be more useful in the long run in helping managers to select sites for protection, reintroduction, or other manipulations.

**LITERATURE CITED**


