CHAPTER 4. METHODS OF ANALYSIS

VARIABLE REDUCTION, TRANSFORMATIONS, AND ORDERING

An Ecological Ordering of Variables

For observational studies with a battery of explanatory variables, James and McCulloch (1990) suggested combining the variables into biologically meaningful groups and then examining all possible subsets with regression analyses. Such an approach is well suited to modeling habitat hierarchically by spatial scale. For my analyses I grouped variables into ecologically meaningful subsets (ecological components) based on similarity of spatial scale and vertical stratum of the forest or aquatic environment (see Dueser and Shugart [1978] and Bingham and Sawyer [1991] for a similar approach). An ecological component was defined as a group of variables that together described a logical class of structural, compositional, or climatic attributes of the forest or stream environment. I derived the ecological components based on biological considerations and the natural relationships I perceived among variables. While such an approach involves some unavoidable anthropocentric bias, it is certainly no greater than the bias imposed by ones' initial choice of variables to measure. On the positive side, this approach has practical application for resource managers, because these ecological components can be directly related to those aspects of forest or stream structure commonly manipulated or altered during such activities as timber harvesting and road building. Significant
variables from each submodel from the ecological component analyses can subsequently be combined in a final analysis to derive composite models that incorporate multiple components and scales. While this approach holds promise for both analytical and management applications, it needs further evaluation by field testing (e.g., different sites from different years), and experimentation.

Variable Reductions

The large number of independent variables (83 or 121) required a strategy for variable reduction in order to perform appropriate statistical analyses. Though concerned with not losing important information, I wanted to eliminate redundant variables and those with many zero values (≥80%) (i.e. those variables representing aspects of the environment that were uncommon, did not vary, or both, on my sample sites—and thus were not useful for estimating amphibian presence or abundance). I first set aside composite variables that were linear combinations of other variables (see Appendix A). I then examined all remaining variables to eliminate or recombine those that contained ≥80% zero values. For example, many counts of logs and their area by decay class contained mostly zero values and were either removed completely (all zeros) or, where possible, combined into composite variables (e.g., counts of large and small decayed logs in two decay classes combined into 'all decayed logs count'). Redundant variables (cover types measured by both the Braun-Blanquet and line transect methods) were
eliminated using correlation analysis; the variable with the highest absolute correlation with animal numbers for each species was retained.

Variable Transformations and Interrelationships

I performed preliminary descriptive analyses to review the distributions of all individual variables. Histograms, normal score plots, and measures of skewness and kurtosis (SAS 1990), were used to assess the normality of distributions, and deviations were corrected by appropriate transformations (log, square root, or arcsine [Sokal and Rolf 1981]). Bivariate scattergrams and correlation coefficients were evaluated to discern relationships among variables. In all multivariate statistical analyses I assumed that univariate normality inferred multivariate normality (Neff and Marcus 1980).

Sample Size Considerations

Block et al. (1987) addressed the problem of observer differences when estimating habitat variables and noted that sample size requirements were generally greater for visually estimated variables as opposed to those measured directly. The majority of variables used in these studies were measured directly, however, I did use some estimated variables (Appendix A). For the two smaller data sets (torrent and Del Norte salamanders), I performed stability analyses (Morrison 1984, Block et al. 1987) to assure that the sample sizes of estimated variables were adequate for subsequent analyses.
STATISTICAL ANALYSES

Univariate Analyses

In addition to the preliminary descriptive analyses described above, in several instances I used univariate statistical methods: (1) as a second tier of analysis, following multivariate applications, to examine relationships with single variables of high significance; or (2) to examine niche relationships where multivariate techniques were not applicable. In the latter case, my samples of the adult tailed frog were insufficient for useful analyses so I included descriptive statistics for selected habitat variables. These applications included both simple descriptive statistics and Pearson product-moment correlation analyses.

Multivariate Analyses

Rotenberry (1986) and Rice et al. (1986) advocated a combined statistical approach--regression analysis plus discriminant analysis--when conducting studies of habitat relationships over a large scale and involving heterogeneous habitat types. Such an approach has the advantage of allowing for cross-validation between methods. Regression analysis explores how the numbers of individuals vary as a particular feature of the habitat varies. A limitation to the regression approach is that factors other than habitat can cause variation in numbers along a habitat gradient thus complicating
interpretation. Discriminant analysis compares the variation in average habitat between sites with and without the species of interest, regardless of differences in population size. This approach addresses the question of what range of attributes of the habitat can provide suitable conditions for the species to exist. Used together in this complimentary way these techniques can reveal aspects of the habitat that may be critical, and therefore possibly limiting for a species, as well as indicate those aspects of habitat which might be managed to maintain or increase animal numbers to insure stable populations.

The overall goal of my analyses was to find those variables that were most strongly associated with variation in animal numbers. Aside from basic ecological information, this approach also provides habitat correlates useful for modeling habitat suitability and predicting the impact of proposed habitat alterations on these species.

Following variable reductions, transformations, and the ordering of variables by spatial scale and ecological components, I performed two types of multivariate statistical analysis: (1) two-group (sites with and without the species of interest) discriminant analyses of habitat variables, arranged in 11 ecological components, and (2), regressions (all-possible-subsets) of the same variables and components, with numbers of each species of interest as the dependent variable.

**Discriminant Analyses of Ecological Components**--Because differences in habitat parameters may not always explain variations in animal
numbers between sites (cf. Van Horn 1983), I first analyzed simple occurrence before exploring variation in animal numbers relative to variation in my independent variables. I used the ecological components in separate discriminant analyses (DA) (SAS 1990) to examine the discriminatory power of each group of variables to discern the presence or absence of each species. This approach resulted in sets of variables that described differences between sites with and without each species of interest.

For each of the four species, a two-group DA (DA I) was run for each ecological component using a stepwise procedure to select variables. Variables were entered into models if their P values for the partial F statistic were \( \leq 0.10 \). For model-building, I chose a moderate significance level (\( P \leq 0.10 \)) permitting more variables to enter my models thus providing the best discrimination power given the limits of my sample size (Costanza and Afifi 1979). In addition, this more moderate level reduces the chances of type II errors (accepting a false null hypothesis). In applied situations related to management of ecological systems, making a type II error is often more tangible and costly than making a type I error because often irreversible management decisions are made on the false premise of no effect (Toft and Shea 1983). This moderate P level also provided a criterion more appropriate for detection of ecological trends (Toft and Shea 1983, Toft 1991).

Once the variables were selected, a linear discriminant function was determined based on those variables. I set an overall significance level for this model at \( P \leq 0.05 \) (model F statistic).
In addition to the assumption of multivariate normality required for statistical inferences, DA also assumes homogeneity of among-groups variance-covariance matrices (Neff and Marcus 1980). Though data are often transformed to correct for lack of homogeneity in order to satisfy assumptions for certain tests, the heterogeneity itself may be of biological interest. For example, it is possible that differences in covariance structure may be a better discriminator between populations than a difference in means (Corruccini 1975). When the variance-covariance matrices are heterogeneous among groups, quadratic discriminant functions may be found for discrimination between any two groups, but they are often difficult or impossible to interpret (Neff and Marcus 1980). I tested for heterogeneity among variance-covariance matrices using Bartlett's modification of the likelihood ratio test (with \( P \leq 0.05 \); SAS 1990) and compared results of classification success for both linear and quadratic functions where appropriate. However, I present all results in terms of linear functions because the quadratic functions yielded the same or only marginally better classification results over the linear functions, linear functions are easier to interpret, and the stepwise technique used to choose variables was linear. Standardized structure coefficients are presented as an indication of the relative contribution of each variable chosen by stepwise DA to the canonical discriminant function (Rencher 1992).

In the case of the tailed frog (*Ascaphus truei*) and the Pacific giant salamander (*Dicamptodon tenebrosus*) the number of sites sampled was sufficient to permit testing of the results of the initial
discriminant analyses by performing a second set of analyses with test data sets removed for subsequent evaluation of results (DA II). The first analysis detailed above (DA I), using the entire data set, consisted of separate DA's of each ecological component and a final DA using those variables derived at the first stage to build a composite model of the niche of each species across spatial scales. The second analysis (DA II) consisted of ten DA runs of each of the eleven ecological components, with random subsets of data removed for model testing. DA II was intended to evaluate the stability of the models from DA I, given the known instability inherent in automatic stepwise selection procedures (see James and McCulloch 1990 and cites therein). Each component model run in DA II had a unique random removal (with replacement) of 25% of the data which were used in subsequent classification tests of the functions derived with the remaining 75% of the data. The final composite DA under DA II consisted of ten additional random subset runs, using those variables derived at the first stage that were selected at least 70% of the time in the ten component model runs.

**Evaluation of Classification Success for Discriminant Models**--I employed two methods to evaluate the classification success of each model in DA I: (1) a jackknife procedure, and (2) a resubstitution procedure (SAS 1990). Cohen's Kappa (Titus et al. 1984) was computed for each test to indicate the classification success compared with chance. For DA II the eleven component models and the final composite models are evaluated using the mean test data and resubstitution results
and mean Kappa statistics derived from the ten runs. The number of times (out of ten runs) that each component yielded a significant model (based on Cohen's Kappa) is also presented. For all classification tests of discriminant models the level for acceptable performance was set at $P \leq 0.05$.

In classification tests based on the discriminant functions I assumed my random systematic site selection procedure yielded a proportion of sites with and without the species of interest that reflects the true proportion. Therefore I adjusted the posterior probabilities of group membership accordingly (priors proportional).

**Multiple Regression Analysis of Ecological Components**—To examine the relationships between variation in animal numbers and variation in habitat variables or other potential niche components, I performed separate all-possible-subsets regression (APS) analyses on each ecological component, and a final analysis on a combination of the best variables from each subset (SAS 1990). In all cases the dependent variable was the number of amphibians at each site. The best model from each ecological component, and the best composite model, were selected on the basis of the highest adjusted $R^2$ value (smallest mean square error) after checking for evidence of multicollinearity (Neter et al. 1989). In the cases of the Del Norte and torrent salamanders sample sizes imposed a second criterion by constraining the total number of variables that could reasonably be considered in a final composite model. Following the guidelines for sample size to variable ratios for
multivariate analyses provided by Johnson (1981), the $R^2$ criterion was applied to a range of models from 7 to 11 variables for the Del Norte salamander, and from 6 to 10 variables for the torrent salamander. I also present the Cp Mallow statistic as an indication of the total mean square error of each model (including both bias and random error components). Significance for the APS analysis was defined as $P \leq .05$ for the component models. However, because I wanted to represent all ecological components in the final APS, all variables from the best model for each ecological component were used to derive the final model regardless of the $P$ value of the component model.

Residual analysis indicated heteroscedasticity in several of my ecological component models. Moderate heteroscedasticity prevents the estimated regression coefficients from being minimum variance estimators, but given that the other conditions of the regression model are met, they are still unbiased and consistent (Neter et al. 1989). I also noted a pattern of points concentrated in a diagonal line in the lower portion of my 'residuals vs. predicted values' scatterplots, for all ecological components. This pattern resulted from the high number of zero values for my dependent variable (roughly half of the sites in each data set had no captures). I acknowledge that including these sites in the analysis can lead to depressed estimates of animal densities, but such estimates were not my primary goal here. Furthermore, I believe this concern is outweighed by the fact that these zero captures sites provide information critical to my evaluation of habitat associations and other niche attributes. In addition, the APS
with the zero captures sites included also provides results that can be compared directly with the discriminant analyses.

**Interpretation of Analyses**—In evaluating my multivariate results I gave more weight to the results of the DA than the APS because variation in numbers beyond simple presence can be influenced by other factors than habitat quality. I consider the question —what attributes of habitat determine the presence or absence of this species?— to be better addressed by my study design than the question of why animal numbers might vary among sites. Results of the APS analyses were considered secondarily and evaluated against the results of the DA; I sought congruencies between the two analyses in terms of the habitat variables that were significantly related to amphibian presence and abundance. Congruency among statistical models was interpreted as strong confirmation of biological significance. However, I do not consider that the relative statistical rankings of habitat variables derived from these analyses necessarily reflect the relative importance of these variables to the ecology of each of these species.

**MODELING HABITAT RELATIONSHIPS**

My models are intended to formalize and quantify our knowledge and understanding of niche relationships for these amphibians. Secondarily, it is hoped that these models will provide a basis for more accurately predicting distributions and abundances of these species in response to anthropogenic and natural changes in the forest environment.
Wildlife-habitat relationships models have become important tools for the management of viable populations of species in our rapidly shrinking natural ecosystems (e.g., Morrison et al. 1992, Patton 1992). Therefore it is useful to present these results in the form of hierarchical habitat relationships models which have a priori value for managing native forestlands. Such models are important tools allowing the planning of management activities so they have minimal impacts on the habitats and life histories of highly specialized species that are dependent on relatively sensitive, rare, or limited structural components of the forest environment. In addition, data relating these amphibians to the forest successional sequence will allow these models to be used to evaluate and improve existing models such as those in the California Wildlife Habitat Relationships (WHR) database (Airola 1988) for the forest types represented in my data sets. Although there exists an extensive literature on modeling wildlife-habitat relationships (e.g., Verner et al. 1986, Patton 1992, Morrison et al. 1992), these efforts are highly biased toward birds and mammals. The related niche literature provides examples from a broader range of taxa (Vandermeer 1972, Colwell and Fuentes 1975), however even here amphibians, are relatively under-represented (see Scott 1982, Toft 1985 and cites therein). In addition, few authors have attempted the approach of integrating several levels of autecological detail, using hierarchical spatial scales, with data derived by sampling from across a large portion of the range of a given species.