

# Predicting the distribution of shrub species in southern California from climate and terrain-derived variables

Franklin, Janet

Department of Geography, San Diego State University, San Diego, CA 92182-4493, USA;  
Fax +1 619 594 4938; E-mail janet.franklin@sdsu.edu

**Abstract.** Generalized additive, generalized linear, and classification tree models were developed to predict the distribution of 20 species of chaparral and coastal sage shrubs within the southwest ecoregion of California. Mapped explanatory variables included bioclimatic attributes related to primary environmental regimes: averages of annual precipitation, minimum temperature of the coldest month, maximum temperature of the warmest month, and topographically-distributed potential solar insolation of the wettest quarter (winter) and of the growing season (spring). Also tested for significance were slope angle (related to soil depth) and the geographic coordinates of each observation. Models were parameterized and evaluated based on species presence/absence data from 906 plots surveyed on National Forest lands. Although all variables were significant in at least one of the species' models, those models based only on the bioclimatic variables predicted species presence with 3 - 26 % error. While error would undoubtedly be greater if the models were evaluated using independent data, results indicate that these models are useful for predictive mapping – for interpolating species distribution data within the ecoregion. All three methods produced models with similar accuracy for a given species; GAMs were useful for exploring the shape of the response functions, GLMs allowed those response functions to be parameterized and their significance tested, and classification trees, while sometimes difficult to interpret, yielded the lowest prediction errors (lower by 3 - 5 %).

**Keywords:** Chaparral; Classification tree; Coastal sage scrub; Direct gradient analysis; Digital terrain model; Generalized additive model; Generalized linear model; Species response curve; Vegetation modeling.

**Abbreviations:** ATM = Atmospheric and Topographic Model; CT = Classification Tree; DEM = Digital Elevation Model; GIS = Geographic Information System; GLM = Generalized Linear Model; GAM = Generalized Additive Model; IPW = Image Processing Workbench; USDA = United States Department of Agriculture; USGS = United States Geological Survey; UTM = Universal Trans-Mercator.

**Nomenclature:** Hickman (1993).

## Introduction

Models that predictively map (Franklin 1995) the distribution of plant species or functional types can be used to interpolate data from costly vegetation field surveys (Austin & Heyligers 1989; Nicholls 1989; Lees & Ritman 1991; Le Duc et al. 1992; Mackey 1994; Cherrill et al. 1995). When they are based on climate variables, these models can be used to make first-order predictions of distribution changes in response to climate change (Austin 1992; Leathwick 1995, 1998; Brzeziecki et al. 1995; Huntley et al. 1995; Kienast et al. 1996; Diaz & Cabido 1997; Guisan et al. 1998; and reviewed by Franklin 1995). Further, these estimates of species' response surfaces (Austin 1985) can also be useful in models of transient ecosystem responses to environmental change when those models require information about the probability of species' occurrence on the landscape or in environmental space. This includes gap models of forest succession (as discussed by Austin et al. 1994a, p. 30), landscape simulation models (for example, He & Mladenoff in press) and dynamic global vegetation models (Steffen et al. 1996).

In order to map predicted plant distributions, spatially-interpolated environmental variables are required (Franklin 1995). It is when those variables are related to direct gradients (*sensu* Austin & Smith 1989) or primary environmental regimes (thermal, radiation, moisture, mineral nutrients; Mackey 1993), and chosen based on functional relationships derived from ecophysiological studies (Box 1981; Woodward 1987), that the response surface models can best be used to address environmental change.

I have developed statistical and decision-tree models predicting the probability of species presence for 20 shrub species that are commonly found in the chaparral and coastal sage scrub communities of southern California. Climate variables (temperature and precipitation) interpolated from climate station data, potential solar radiation estimated across a topographic grid, and a terrain variable related to potential soil moisture, were tested as explanatory variables.

Chaparral, characterized by a continuous canopy of evergreen sclerophyllous shrubs, and coastal sage scrub,

characterized by a more discontinuous cover of drought-deciduous malacophyllous shrubs and subshrubs, are found in the coastal and lower montane regions of southern California, which experience a Mediterranean-type climate regime with wet cool winters and dry warm to hot summers (Hanes 1977; Mooney 1977; Keeley & Keeley 1988). The distribution of the various shrub species has been related to indirect gradients such as (1) elevation (related to both precipitation and winter minimum temperatures); (2) slope aspect (related to solar insolation, and hence potential evapotranspiration); (3) slope angle or topographic position (related to soil depth or texture, and thus potential soil moisture); and (4) substrate and lithology (related to nutrient and moisture availability) (Wells 1962; Kirkpatrick & Hutchinson 1980; Poole & Miller 1981; Westman 1981; Miller et al. 1983; Keeley & Keeley 1988; Davis & Goetz 1990; Gordon & White 1994; previously reviewed in Franklin et al. in press). Westman (1991) modeled the range of several chaparral and coastal sage shrub species in relation to climate variables derived from small-scale climate maps.

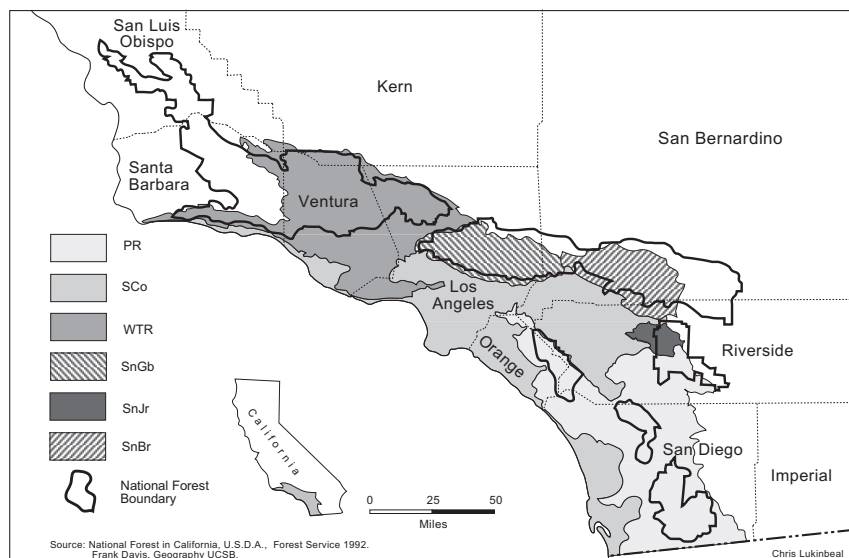
## Materials and Methods

### *Data on species distributions*

Visual estimates of the cover of plant species at 906 locations in three National Forests within California's 'southwest ecoregion' (defined and delineated in Hickman 1993 and Davis et al. 1994; Fig. 1) were

available from the USDA Forest Service. Note that these 'forests' are actually dominated by shrub formations covering 50 - 80 % of their land area, and the 10 - 30 % of the land area covered by forest and woodland is not managed for timber production due to its limited extent and slow growth (Franklin 1996; Franklin & Woodcock 1997). The data were collected in 10-m radius circular plots between 1990 and 1994 for the purpose of defining shrub vegetation types or 'series' (Gordon & White 1994). Of 1215 plots originally collected, these 906 had their locations recorded as points on 1:24 000 topographic maps, and were digitized. While the accuracy with which plots were located on maps is not known, I would estimate it to be within 50 m. The plot locations were subjectively chosen. Sampling was both purposive (to sample different chaparral associations throughout the study area and elevation range) and opportunistic (locations were chosen based on ease of access and are clustered near roads). Species were not necessarily sampled evenly throughout geographical or gradient space. Also, all of these species have ranges that extend beyond the ecoregion, at least into northern Baja California (see Westman 1991; Hickman 1993). This implies that the models described below can be used to interpolate species distributions within the study area, but not to predict range changes nor to examine the shape of the response function describing the realized species niche (Austin et al. 1990).

Data for the 104 most frequently occurring species (of the 330 taxa of flowering plants sampled) were made available to me. Preliminary examination of patterns of



**Fig. 1.** Study area: plots were located within the southernmost National Forests in California (outlined); terrain data were available for the National Forest lands, climate variables (see Fig. 2) for the southwest ecoregion as defined in Hickman (1993) and Davis et al. (1994). Subdivisions of the ecoregion include: WTR = Western Transverse Ranges; SnGb = San Gabriel Mountains; SnBr = San Bernardino Mountains; SnJt = San Jacinto Mountains; PR = Peninsular Ranges; SCo = South Coast.

species abundance (cover) in relation to the environmental variables showed a great amount of scatter, and therefore I converted the data to species presence/absence for modeling. The 20 species modeled were chosen based on their high frequency of occurrence in the sample, their importance as indicators of certain vegetation types defined by Gordon & White (1994) (*Ceanothus tomentosus*, *Xylococcus bicolor*, *Adenostoma sparsifolium*, *Artemisia californica*), or their having been the subject of previous modeling efforts (*Ceanothus greggii*, *Malosma laurina*; Westman 1991).

#### *Mapping the environmental variables*

Climate variables and derived bioclimatic indices represent primary environmental regimes related to the physiological tolerances of organisms (Box 1981, 1996; Woodward 1987; Woodward & Williams 1987; Mackey & Sims 1993; Mackey 1993, 1994). They are measures of, or surrogates for, energy and water inputs, seasonal extreme conditions, and conditions during growing and dormant seasons and have been used to model and predictively map biotic distributions (reviewed in Austin et al. 1994a, p. 11; Franklin 1995; and see also Leathwick 1995, 1998; Brzeziecki et al. 1995; Beerling et al. 1995; Huntley et al. 1995; Guisan et al. 1998). Several approaches to interpolating climate variables on a geographic and topographic grid have been developed and tested (Hutchinson 1984, 1987; Daly et al. 1994; Hutchinson et al. 1996).

J. Michaelsen (unpubl.) provided 1-km gridded estimates of 28 climate variables for the southwest ecoregion: mean minimum and maximum monthly temperature for each month, and annual and trimesterly mean precipitation. The surfaces were interpolated from climate station data using elevation from a digital elevation model. Universal kriging (Cressie 1991) was the interpolation method used in a two-component model. Ordinary kriging was used to model the spatial dependence of residuals from regression models predicting the climate variables from up to four variables: location (latitude, longitude), elevation, and a topographic barrier index to capture the rainshadow effect. Then the regression model was re-estimated using the variogram model of spatial dependence, and this was repeated until the estimates converged (as suggested by Bailey & Gatrell 1995; Michaelsen pers. comm.).

These interpolated values of the climate variables at the vegetation plot locations are estimated with error (0.5 - 2.0 °C for the temperature variables), and this will add unexplained variance to the models described below. Further, the climate variables are interpolated to a resolution (1 km × 1 km) which is much larger than the vegetation plot (0.03 ha). However, because cli-

mate is slowly-varying in space, relative to individual plants, or topography, this scale problem should not introduce additional noise.

Because there is a body of literature and theory linking bioclimatic gradients to plant distributions, and because the models are intended to be predictive rather than exploratory, explanatory variables were chosen *a priori* (Judge et al. 1985, p. 869). Three of the climate variables were selected for modeling: annual precipitation (strongly related to growing season precipitation in this Mediterranean-type climate – precipitation in the dormant season is near zero): ‘ppt.yr’; mean minimum temperature of the coldest month (December): ‘dec.mint’; and mean maximum temperature of the warmest month (July): ‘jul.maxt’ (Fig. 2). The deviation of the coldest temperature from the annual mean (suggested by Leathwick 1995) was also examined, but was strongly correlated with jul.maxt for this data set (Pearson’s  $r = -0.72$ ).

Potential solar radiation (insolation) was estimated using the Atmospheric and Topographic Model (ATM) of solar radiation (Dozier 1980, 1989; Dozier & Frew 1990; Dubayah 1992; Dubayah & Rich 1995) implemented in the Image Processing Workbench (IPW; Frew 1990). This approach has sufficient physical realism to simulate the terrain-distributed surface energy balance, and has been validated using field data (Dubayah 1992; Dubayah & Rich 1995). However, for the purpose of modeling species response surfaces I was primarily interested in the relative radiation balance of different landscape positions (related to potential evapotranspiration; Dubayah 1994). Therefore, I estimated only clear-sky potential incoming shortwave radiation using a number of simplifying assumptions (uniform land surface albedo, constant uniform atmospheric optical properties for the entire model year). Topographic variables (elevation, slope and aspect) used in radiation modeling were derived from USGS 7.5’ 30 m × 30 m resolution DEMs. Incoming solar radiation was calculated for each grid cell in the DEM, for a single day each month, and then multiplied by the number of days as an estimate of monthly potential radiation. These values were summed to estimate quarterly and annual solar radiation.

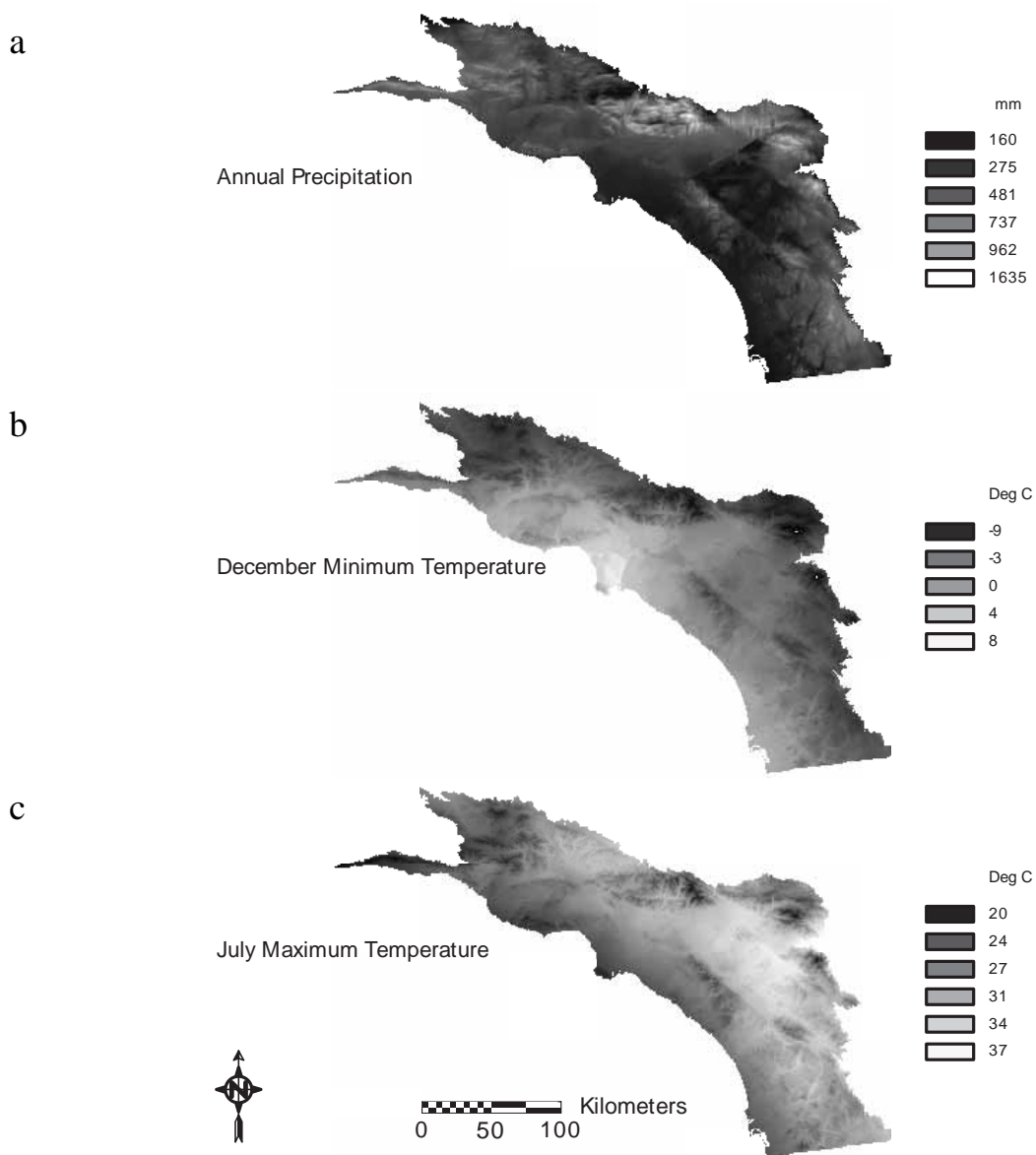
Quarterly estimates of solar radiation for the growing season (spring: Mar-May), and driest quarter (summer: Jun-Aug) were highly correlated ( $r = 0.88$ ), and therefore only winter (Dec-Feb) and spring values were used in modeling (‘psi.win’, ‘psi.spr’). In southern California, the difference in energy balance and potential evapotranspiration between north- and south-facing slopes is greatest during the spring and summer (Miller et al. 1983). However, winter potential insolation showed a strong empirical correlation with vegetation patterns, probably because illumination differences between

north- and south-facing slopes are most pronounced due to low sun angle (Davis & Goetz 1990; McCullough 1995; Franklin et al. in press).

Errors commonly found in DEMs (related to the method used to derive the elevation grid) are amplified when first- and second-order difference operations are applied to derive slope, aspect and other terrain variables (Zevenberger & Thorne 1987; Skidmore 1989; Weibel & Heller 1991; Brown & Bara 1994; reviewed in Franklin et al. in press). In this dataset, the correlation between elevation recorded with the field data (from topographic maps) and the DEM-derived elevation for the plot locations was  $r = 0.99$ , while the correlations for

field- versus DEM-estimated slope (angle or gradient;  $r = 0.48$ ) and aspect ( $r = 0.46$ ) were much lower (see also Davis & Goetz 1990). This had an unknown effect on the estimates of potential solar radiation. However, differences between the values may be due to scale differences in the measurements (discussed by Franklin et al. in press).

Additional mapped variables, related to environmental gradients, have been found important in vegetation modeling (reviewed by Franklin 1995) including soil type, landform, geology, and terrain variables related to soil depth and texture (slope angle, slope curvature, hillslope position) and potential soil moisture (drainage



**Fig. 2.** Climate variables used in the models: **a.** Mean annual precipitation; **b.** Mean minimum temperature of the coldest month (December); **c.** Mean maximum temperature of the warmest month (July).

**Table 1.** Bioclimatic, topographic and spatial variables used as explanatory environmental variables in the models.

Variable name	Units	Description
ppt.yr	mm	average annual precipitation
dec.mint	°C	average minimum daily temperature of the coldest month (December)
jul.maxt	°C	average maximum daily temperature of the warmest month (July)
psi.spr	Wm <sup>-2</sup>	topographically-distributed potential solar insolation for the spring quarter (Mar-May)
psi.win	Wm <sup>-2</sup>	topographically-distributed potential solar insolation for the winter quarter (Dec-Feb)
slope.deg	°	slope angle or gradient
utm.n	m	geographic coordinate — UTM northing
utm.e	m	geographic coordinate — UTM easting

basin position, the topographic moisture index or TMI described by Moore et al. 1991). Digital maps of substrate or lithology were not available for the entire study area at an appropriate scale. Slope aspect is strongly related to, and was used to model, potential solar radiation (Dubayah 1994) and therefore was not included separately. Other terrain variables such as slope curvature and TMI were evaluated but not used, due to the errors the DEM described above. The only additional variable that was tested was slope which was calculated using the IPW software and a second order finite difference method (Skidmore 1989; Frew 1990). Explanatory variables used in the vegetation models are summarized in Table 1.

Finally, many statistical models of species response surfaces have ignored the potential for untreated spatial dependence (autocorrelation) in the response variable to inflate the estimated explanatory power of these models. Some studies have tested for spatial autocorrelation in the explanatory variables, and then used systematic sampling at a distance interval for which spatial randomness could be assumed, to generate a sample that met the independence assumption of a generalized linear model (Davis & Goetz 1990). If the spatial patterning of species distributions is caused entirely by spatial patterning of the explanatory environmental variables, then it may not be necessary to parameterize the spatial structure of the dependent variable in the model (Smith 1994). However, if there is spatial structure in the species data that is not shared by the environmental data, it may reflect biotic processes such as dispersal, predation and competition, as has been discussed at length in the literature (Pielou 1977; Ripley 1981). While characterizing the spatial structure of the species distribution may produce more accurate predictions of its distribution within the study area, removing spatial structure from the model may make it sufficiently general to extrapolate outside the study area (Harvey 1996).

A technique that has been used is to generate a trend surface model for the dependent variable, predicting its response using a quadratic or cubic trend surface regression based on the geographic coordinates ( $x$ ,  $y$ ) of the observations, and combine this with a model based on environmental variables using Bayesian or other

techniques (Pereira & Itami 1991; Le Duc et al. 1992; see also Bocard et al. 1992). Another approach is to use an estimate of clustering as an additional explanatory variable in a statistical model (Smith 1994; Leathwick 1998). I developed models both with and without geographic coordinates as explanatory variables. This addresses geographic patterning but not spatial dependence or clustering in the dataset.

#### *Methods of analysis*

Three modeling frameworks being used with increasing frequency to describe species' response surfaces are generalized linear models (GLM, McCullagh & Nelder 1989; Crawley 1993), generalized additive models (GAM, Hastie & Tibshirani 1990; Yee & Mitchell 1991) and decision trees or classification and regression trees (Breiman et al. 1984; Clark & Pregibon 1992). These modeling approaches were reviewed in Franklin (1995) and Austin et al. (1994a; see also Leathwick 1995, 1998; Bio et al. 1998). Austin & Meyers (1995), and Austin et al. (1995) compared these and several other modeling approaches. In the present study, I used species presence/absence as the dependent variable, and so the logistic link function and binomial error term were used in the GAMs and GLMs, and classification trees (CT) were developed to predict species presence or absence.

GAMs were first parameterized using all environmental and spatial variables, and smoothing splines, to assess the validity of the linear model for each bivariate relationship, and to explore the shape of the response curve and strength of the relationship between each species and explanatory variable. Then, GLMs were parameterized using all variables and using response functions suggested by the GAM. The following restrictions were imposed: all response functions were modeled as either linear, quadratic (approximating a unimodal or symmetric Gaussian response curve), or a higher-order polynomial. For simplicity, and because the sample did not span the entire species ranges, skewed quadratic responses were not parameterized, although they are frequently cited in the literature (Minchin 1987; Austin et al 1994b; but see Oksanen 1997 and Austin & Nicholls

1997), nor were piecewise linear (monotonic) relationships (Bayes & Mackey 1991; Mackey 1993). Interaction terms were initially evaluated for significance in the GLMs, and although they were occasionally significant, they were not included in the models presented below. Variables in the GLMs were tested for significance, and deleted if not significant (backward elimination), and the remaining variables again tested for significance.

Finally, classification trees were developed for each species. A cut-off of five observations per terminal node was used, beyond which no splitting of the dataset occurred. Cross-validation was used to determine the optimum tree size yielding the most robust predictions (Clark & Pregibon 1992; Venables & Ripley 1994; Michaelsen et al. 1994). Cross-validation suggested that deviance was minimized at about 10 - 24 terminal nodes for many of the species models; therefore, for expediency, all trees were pruned to the 22 'best' nodes based on a cost-complexity measure (deleting those 'branches' that reduce deviance the least). Then the tree was pruned interactively to the smallest size possible without increasing model error (by aggregating terminal nodes that gave the same prediction). This generally resulted in models that used a subset of the explanatory variables. Statistical analyses were carried out using the S-Plus software version 3.3 for Windows (Anon. 1995).

It has been noted that when observations from outside the biogeographic range of the species are used to develop a model, this can inflate the amount of deviance explained by the model (Austin et al. 1995, p. 18; see also Guisan et al. 1998). Austin et al. (1995) have recommended that these 'structural zeroes' or 'naughty noughts' (Austin & McKenzie 1988) be removed from the dataset. They suggest deleting observations that are beyond the last positive observation by greater than 1% of the sample, sorted by an important environmental variable. I included the last positive observation  $\pm 10\%$  of the sample, and so the fit of my models may still be overestimated somewhat. For example, there were roughly 1000 observations (906), and if they are sorted by mean minimum temperature of the coldest month, and the first occurrence of a species is in the 200th observation and the last in the 500th, I selected the 100th-600th observations, sorted thus, and use that subset of the data to develop models for that species. I sorted, iteratively, for up to three climatic variables.

#### Model evaluation

In all three of these modeling frameworks, the model fit and significance of the variables are evaluated using the residual deviance (analogous to the residual sum of

**Table 2.** Summary of the shape of response functions for each environmental variable (Table 1), interpreted from smoothing functions fit in GAMs, and functions used to parameterize GLMs. Number of species (from among 15 chaparral or five coastal sage species) showing each type of response curve for GAMs, and number of species for which the variable was significant in GLMs ( $p < 0.05$ ).

Response function	ppt.yr	dec.mint	jul.maxt	psi.spr	psi.win	slope.deg	utm.n	utm.e
<b>Chaparral (<math>n = 15</math>)</b>								
<i>GAMs</i>								
Linear	0	0	1	1	1	0	0	1
Piecewise linear	6	6	8	7	7	3	7	6
Skewed quadratic	5	9	2	0	2	0	1	2
Bimodal	0	0	1	1	1	3	2	2
Higher-order polynomial	1	0	3	0	1	1	4	3
No apparent relationship	3	0	0	6	3	8	1	1
<i>GLMs</i>								
Significant	9	13	11	7	9	2	13	8
Linear	4	5	3	7	9	2	7	3
Quadratic	5	8	7	0	0	0	3	4
Higher-order polynomial	0	0	1	0	0	0	3	1
<b>Coastal sage (<math>n = 5</math>)</b>								
<i>GAMs</i>								
Linear	0	0	0	0	2	0	0	0
Piecewise linear	3	5	1	4	3	1	4	2
Skewed quadratic	0	0	3	0	0	2	0	1
Bimodal	2	0	0	0	0	0	0	0
Higher-order polynomial	0	0	1	0	0	0	1	1
No apparent relationship	0	0	0	1	0	2	0	1
<i>GLMs</i>								
Significant	5	5	3	2	4	3	3	2
Linear	2	5	0	2	3	1	1	0
Quadratic	3	0	3	0	1	2	1	1
Higher-order polynomial	0	0	0	0	0	0	1	1

squares in the linear model). This is not as intuitively meaningful as the 'proportion of variance explained' from a linear model, but residual deviances can be compared with analysis of variance using a  $\chi^2$ -test. However, because the CT-model is restricted to a normal error function (Austin et al. 1995, p. 74), rather than one more appropriate for binary data, comparisons of residual deviance between trees and the other types of models are invalid. The models were also assessed by examining the number of observations (presence/absence) correctly predicted by each model, as well as the proportion of omission (type I) and commission (type II) errors – predicting the species to be absent when it is present, and *vice versa*. Both can be important types of error in species distribution models. Further, because these models predict the probability,  $p$ , of the response variable being true, in this case of the species being present – which I will refer to as  $p(P)$  – commission and omission errors should be examined as a function of  $p$  for each model. An optimum threshold value of  $p$  should be chosen that minimizes omission and commission errors (Pereira & Itami 1991; Guisan et al. 1998 used the Cohen's  $k$  for calibrating this threshold probability value). This is important because when a species is rare in the sample, a 'model' that predicted it to be absent everywhere would have high overall accuracy (but high omission error), and *vice versa*. Although the data used to estimate model fit and prediction accuracy are the same as those used to develop the models, the diagnostics are still useful for comparisons among models.

## Results

### *GAMs and GLMs: Variable selection and response curves*

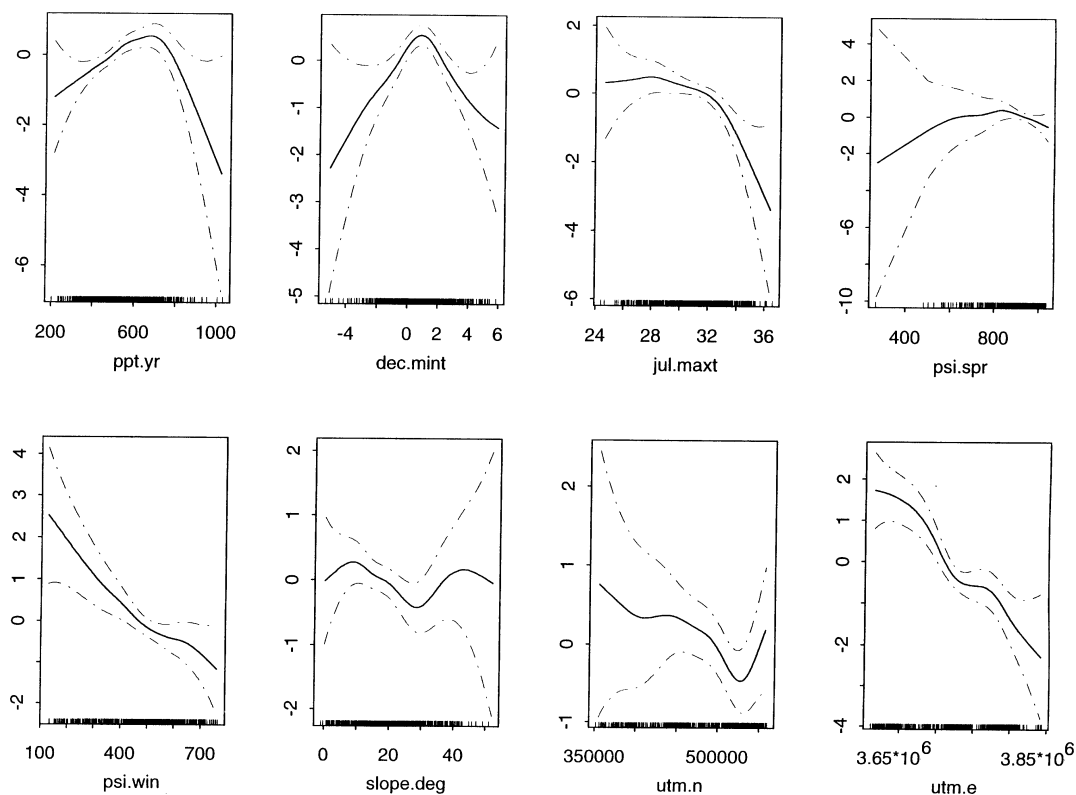
GAMs parameterized for the 20 species suggested skewed quadratic or piecewise linear responses to annual precipitation (ppt.yr) and winter monthly mini-

mum temperature (dec.mint) for the chaparral species (Table 2, Fig. 3), but mainly piecewise linear responses for the coastal sage species (Table 2, Fig. 4). The latter are reaching their upper elevational limit within the study area (the National Forest lands do not extend to the coast, Fig. 1). The pattern of response to summer monthly maximum temperature (jul.maxt) was reversed – usually piecewise linear for chaparral species and skewed quadratic for coastal sage species (Table 2). This may be because highest jul.maxt occurs in the interior valleys and middle elevations of the study area, where coastal sage species are found (Fig. 2). The response functions for geographic coordinates ranged from linear to fourth-order polynomial. Plots of the smoothing functions for the GAMs suggested no apparent relationship to spring potential solar insolation (psi.spr) for about a third of the species, or to winter insolation (psi.win) in about a quarter of the cases (Table 2). These plots usually suggested a stronger relationship for psi.win than for psi.spr (Figs. 3, 4). Slope angle (slope.deg) had no apparent relationship to species response for half of the species (Table 2, Fig. 3).

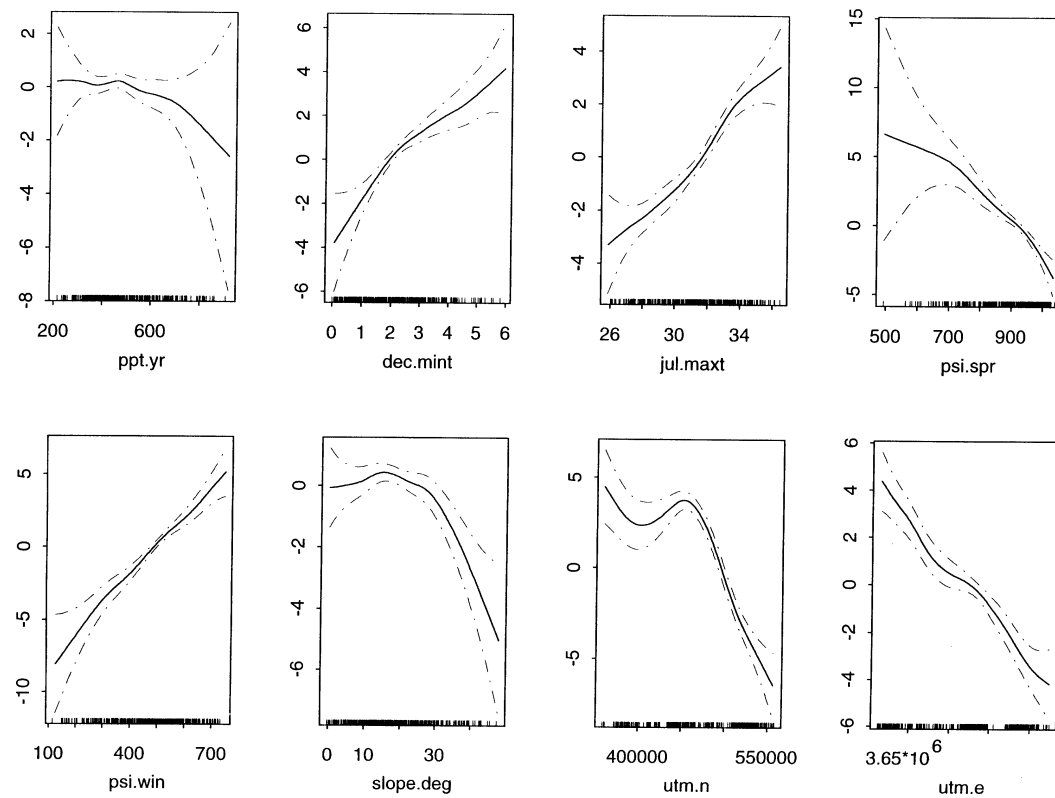
GLMs parameterized using linear terms and second-order polynomials (higher-order polynomials only for geographic coordinates) showed that the precipitation, temperature and geographic variables, especially dec.mint, were retained in the models most frequently (significant at  $p < 0.05$ , Table 2). These first five variables, when significant, always explained the most deviance in the GLMs, although the exact magnitude and order of importance of the variables varied greatly among the species. Potential winter insolation (psi.win) was also significant in more than half of the chaparral models and most of the coastal sage models, but always explained less deviance than the previous five variables. Spring potential insolation (psi.spr) was significant in less than half of the models. Slope (slope.deg) was significant for only five of the species (Table 2, Fig. 4).

**Table 3.** Frequency with which the five bioclimatic and two spatial variables (Table 1) were selected in the tree model for the chaparral species ( $n = 15$ ) and coastal sage scrub species ( $n = 5$ ). The ranking indicates where in the tree model the variable was first selected, e.g., a variable ranked 1 - 3 was used in splits 1 - 3 of the tree (explained the highest amount of deviance).

Species	ppt.yr	dec.mint	jul.maxt	psi.spr	psi.win	utm.n	utm.e
<b>Chaparral (<math>n = 15</math>)</b>							
Frequency	9	14	13	10	13	13	14
Frequency ranked 1-3	4	11	8	1	5	7	9
Frequency ranked 4-5	4	2	4	7	6	4	2
Frequency ranked 6-7	1	1	1	2	2	2	3
<b>Coastal sage scrub (<math>n = 5</math>)</b>							
Frequency	4	5	5	4	5	4	5
Frequency ranked 1-3	2	4	3	1	1	0	4
Frequency ranked 4-5	1	1	2	1	3	3	0
Frequency ranked 6-7	1	0	0	2	1	1	1

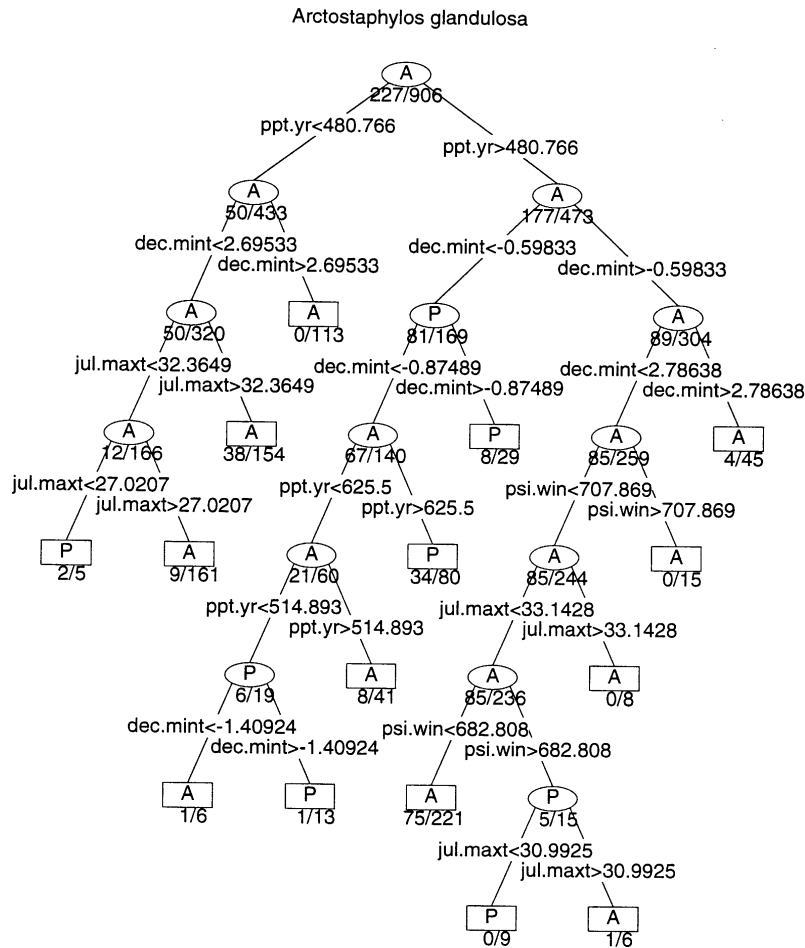


**Fig. 3.** GAM smoothing functions with 95% confidence intervals (dashed lines) for the eight environmental variables described in Table 1, for *Quercus berberidifolia*. The y-axis in each plot is the smoothing function applied to the variable.



**Fig. 4.** GAM smoothing functions with 95% confidence intervals (dashed lines) for the eight environmental variables described in Table 1, for *Artemisia californica*. See caption Fig. 3.





**Fig. 5.** Classification tree model for *Arctostaphylos glandulosa* based on bioclimatic variables (see Table 6). Abbreviations used for variables are given in Table 1. The ellipses represent internal nodes and rectangles terminal nodes of the pruned tree. Symbols P = Present, A = Absent are the models' prediction of the species' presence/absence at each node. The ratio below each node is the proportion of observations misclassified at that node.

### Classification trees: Variable selection

Based on the variable selection for the GLMs, and preliminary testing of classification tree models, slope was not included as a variable in the tree models presented here. CT-models developed using the remaining seven variables, and pruned as described above (number of terminal nodes ranged from 9 - 18), selected the temperature and geographical variables and psi.win for almost all species, while ppt.yr and psi.spr were selected in about two thirds of the cases (Table 3). Further, variables are selected in a CT-model in order of how much deviance they explain; dec.mint, jul.maxt and utm.e (related to distance from the coast) were most frequently selected first, second or third, while psi.spr and psi.win were most frequently selected fourth or fifth (Table 3). This not only mirrors the pattern found in the GLMs, it suggests that the tree model may be capturing a hierarchical interaction between the climatic and topographically-distributed insolation variables that was not captured in the GLMs where insolation variables were significant less frequently. For example, the CT-model based on bioclimatic variables

for *Arctostaphylos glandulosa* (Fig. 5) can be interpreted as follows: when precipitation is low (left main branch), the species is present only if summer maximum temperature is also low. If precipitation is not so low (right main branch), and if winter minimum temperature is not too high, and if psi.win is high (but not too high), and if summer maximum temperature is moderate, then the species is present. In the CT-model for *Ceanothus tomentosus* based on five bioclimatic variables (tree not shown), the species is predicted to be present at lower psi.win (< 489 W/m<sup>2</sup>) if summer maximum temperature is high (29 - 32 °C), while it is predicted to be present at higher psi.win (up to 536 W/m<sup>2</sup>) if summer maximum temperature is lower (< 27 °C).

CT-models were then developed excluding the geographic variables. The number of nodes ranged from 9 - 21. The five remaining variables were retained more frequently in the models. The temperature variables were most frequently selected first or second, ppt.yr and psi.win third or fourth, and psi.spr last (Table 4).

Further, the 'response functions' as they are approximated in the tree models by the selection of threshold

**Table 4.** Frequency with which the five bioclimatic variables only were selected in the tree model for the chaparral species ( $n = 15$ ) and coastal sage scrub species ( $n = 5$ ). See Table 3 caption.

Species	ppt.yr	dec.mint	jul.maxt	psi.spr	psi.win
<b>Chaparral (<math>n = 15</math>)</b>					
Frequency	15	15	15	12	14
Frequency ranked 1 - 2	6	10	11	0	3
Frequency ranked 3 - 4	9	4	3	4	10
Frequency ranked 5	0	1	1	8	1
<b>Coastal sage scrub (<math>n = 5</math>)</b>					
Frequency	5	5	5	3	5
Frequency ranked 1 - 2	2	4	2	0	2
Frequency ranked 3 - 4	3	1	3	1	1
Frequency ranked 5	0	0	0	2	2

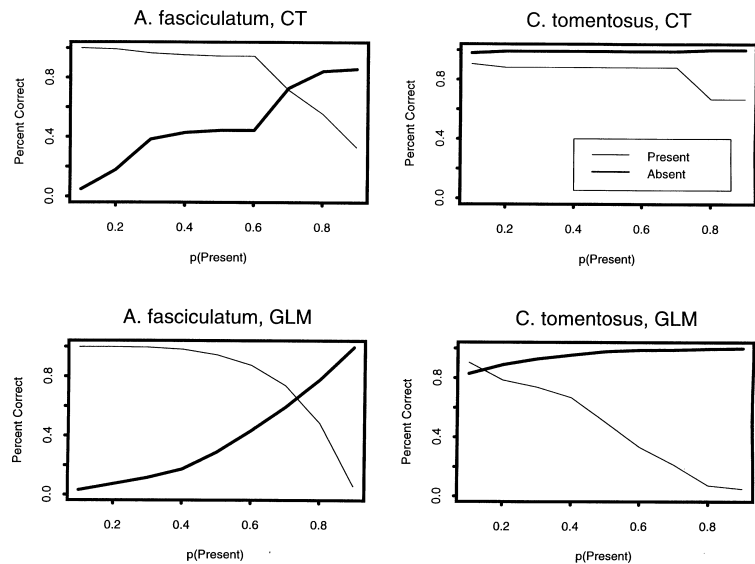
values of variables, were generally ecologically rational (*sensu* Austin et al. 1995, p. 77) based on observations of the species' abundance versus environmental variables (not shown). However, interactions between variables modeled by their nested selection within the tree are admittedly difficult to evaluate for their ecological rationality, even for these relatively simple or parsimonious trees (discussed in Austin et al. 1995).

#### All models: Predictions and error

When GAMs developed using all (eight) variables are compared to the GLMs that only retain significant variables (three-seven), the GAMs are almost always significantly better ( $p < 0.01$ ) based on the  $\chi^2$ -test (Table 5) – they explain more deviance relative to the number of degrees of freedom lost to estimate the smoothing functions. The GAMs and GLMs appeared to explain a modest proportion of the deviance (high residual deviance; Table 5), but this is typical for this type of study and is related to the binary nature of the dependent variable (discussed in Bio et al. 1998). Perhaps a more easily understood measure of their fit is the prediction error (proportion of observations incorrectly predicted present and absent, Table 5). This ranges from 2% to 27% for the GAMs and 5% to 30% for the GLMs. On average the GLM had 2% higher prediction error than the GAM for the same species for reasons discussed above (range 0.02%-5%). Prediction accuracy was generally higher for those species that were relatively rare in the sample and which tended to be restricted to certain portions of the environmental gradients (for example, *Xylococcus bicolor*, *Malosma laurina*, *Ceanothus tomentosus*). Recall that these prediction errors may mask very high omission or commission errors (as shown below) and should be mainly used for comparisons among the models.

**Table 5.** Summary of generalized additive models and generalized linear models for the 20 species analyzed. n obs = number of observation (plots) used in species' model; n pres = number of those plots in which the species was present; Null Dev = null deviance; Resid Dev = residual deviance (unexplained by model); Resid DF = residual degrees of freedom; n var = number of variables used in the model (Table 2) – note that the number of variables in the GAMs is always 8; Error = proportion of observations incorrectly predicted when probability threshold,  $p(P)$ , = 0.5; pr  $\chi^2$  = probability that the GAM and GLM are not significantly different based on the  $\chi^2$ -test.

Species	n obs	n Pres	Null Dev	GAM				Error	GLM				pr $\chi^2$
				Resid Dev	Resid DF	n var	Resid Dev		Resid DF	n var	Error		
Chaparral													
<i>Adenostoma fasciculatum</i>	906	856	1087.97	905.32	873.31	8	0.23	953.22	896	5	0.25	0.0015	
<i>Adenostoma sparsifolium</i>	535	100	515.44	173.89	503.57	8	0.07	291.98	528	6	0.12	0.0000	
<i>Arctostaphylos glandulosa</i>	906	296	1020.05	807.74	873.72	8	0.22	885.91	900	3	0.26	0.0000	
<i>Arctostaphylos glauca</i>	906	295	997.54	737.04	873.64	8	0.19	765.03	892	7	0.20	0.0695	
<i>Ceanothus crassifolius</i>	906	194	803.41	461.15	873.99	8	0.11	556.76	896	5	0.15	0.0000	
<i>Ceanothus greggii-perplexans</i>	841	200	883.67	461.62	808.99	8	0.13	528.21	830	6	0.15	0.0000	
<i>Ceanothus leucodermis</i>	906	230	826.00	591.85	873.94	8	0.15	649.77	892	7	0.16	0.0000	
<i>Ceanothus tomentosus</i>	375	43	263.00	89.24	342.71	8	0.05	146.09	368	5	0.07	0.0003	
<i>Cercocarpus betuloides</i>	906	301	1039.30	848.28	873.53	8	0.24	892.27	897	5	0.26	0.0063	
<i>Heteromeles arbutifolia</i>	868	122	599.50	439.55	835.30	8	0.10	493.95	859	6	0.11	0.0003	
<i>Quercus berberidifolia</i>	906	417	1149.16	835.88	873.86	8	0.22	904.90	894	7	0.25	0.0000	
<i>Quercus wislizenii</i>	906	193	612.43	380.20	873.76	8	0.09	418.96	899	5	0.09	0.0426	
<i>Rhus ovata</i>	906	170	806.69	678.28	873.85	8	0.15	738.67	899	5	0.16	0.0000	
<i>Xylococcus bicolor</i>	322	25	175.79	42.36	290.02	8	0.02	87.15	314	5	0.05	0.0060	
<i>Yucca whipplei</i>	906	328	1091.56	992.91	873.19	8	0.27	1053.68	899	4	0.29	0.0001	
Coastal sage scrub													
<i>Artemisia californica</i>	575	127	572.70	265.87	542.74	8	0.09	358.04	566	5	0.14	0.0000	
<i>Eriogonum fasciculatum</i>	906	387	1176.72	990.40	873.29	8	0.27	1025.7	895	7	0.30	0.0330	
<i>Malosma laurina</i>	575	47	279.90	134.73	542.55	8	0.05	175.17	566	5	0.07	0.0159	
<i>Salvia apiana</i>	906	185	838.56	649.05	873.43	8	0.16	710.46	899	5	0.18	0.0000	
<i>Salvia mellifera</i>	806	175	705.55	416.27	773.67	8	0.13	467.88	796	6	0.13	0.0004	



**Fig. 6.** Prediction accuracy for species presence and absence, as a function of the  $p$ -value (probability of species presence) of the GLM (Tables 2 and 5) and CT models (based on all seven variables, Tables 3 and 6), for *Adenostoma fasciculatum* and *Ceanothus tomentosus*.

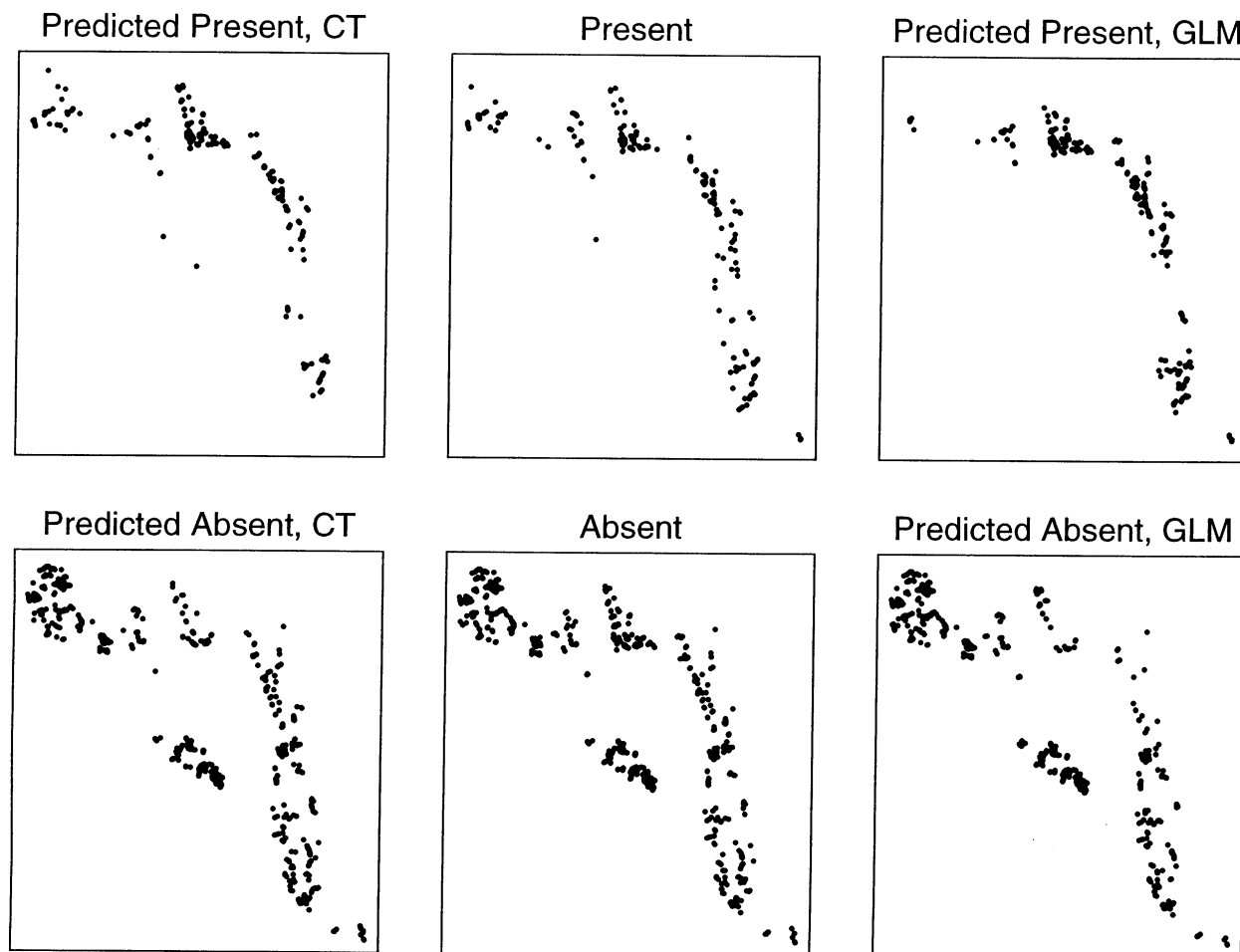
The CT-models based on up to seven bioclimatic and spatial variables explained more deviance than the GAMs or on the GLMs based on all eight original variables (Table 6). Tree models based on the five bioclimatic variables only also explained more deviance than almost all of the GLMs which included spatial variables (19 of 20), and many of the GAMs which included all eight variables (11 of 20). Recall, it is not possible to test the significance of these differences.

Prediction error was 2 - 25 % for the tree models with spatial variables, and 3 - 25 % for those without spatial variables – on average 4.5 % lower than for the GLMs (range 0.03 - 8 %).

Recall that each of these models predicts a probability of the binary dependent variable (based on the logit link function, or the proportion of class membership in the terminal node in the case of CTs). When omission and commission error are calculated for the models as a

**Table 6.** Summary of classification tree models for each species based seven variables (Table 3), versus five bioclimatic variables only (Table 4); node = the number of terminal nodes in the final model (pruned from 22 nodes as described in the text); other columns are as described in Table 5.

Species	All Variables (7)					Bioclimatic variables (5)				
	node	Resid Dev	Resid DF	n var	Error	node	Resid Dev	Resid DF	n var	Error
<b>Chaparral</b>										
<i>Adenostoma fasciculatum</i>	15	839.4	891	6	0.20	16	847.4	890	5	0.21
<i>Adenostoma sparsifolium</i>	14	120.0	521	6	0.04	14	217.3	521	4	0.07
<i>Arctostaphylos glandulosa</i>	16	724.0	890	7	0.22	15	760.0	890	4	0.20
<i>Arctostaphylos glauca</i>	16	669.7	890	5	0.15	17	699.6	889	5	0.18
<i>Ceanothus crassifolius</i>	18	373.4	888	6	0.09	19	455.1	887	5	0.13
<i>Ceanothus greggii perplexans</i>	18	334.3	823	5	0.10	12	596.4	829	4	0.15
<i>Ceanothus leucodermis</i>	17	512.3	889	6	0.11	17	565.0	889	5	0.13
<i>Ceanothus tomentosus</i>	11	46.59	364	4	0.02	16	71.5	359	5	0.04
<i>Cercocarpus betuloides</i>	17	743.5	889	6	0.18	18	791.7	888	5	0.18
<i>Heteromeles arbutifolia</i>	17	356.2	851	6	0.08	16	397.7	852	4	0.09
<i>Quercus berberidifolia</i>	18	766.0	888	6	0.19	21	846.0	885	5	0.21
<i>Quercus wislizenii</i>	13	373.9	893	6	0.09	13	386.4	893	5	0.08
<i>Rhus ovata</i>	17	588.0	889	6	0.13	18	604.0	888	5	0.13
<i>Xylococcus bicolor</i>	9	41.8	313	5	0.03	15	53.0	307	5	0.03
<i>Yucca whipplei</i>	14	927.0	892	6	0.25	16	927.0	892	5	0.24
<b>Coastal sage scrub</b>										
<i>Artemisia californica</i>	16	238.5	559	7	0.07	13	291.6	562	5	0.09
<i>Eriogonum fasciculatum</i>	17	907.9	889	7	0.23	14	957.7	892	5	0.25
<i>Malosma laurina</i>	13	143.7	562	6	0.04	13	150.0	562	5	0.04
<i>Salvia apiana</i>	14	616.7	892	6	0.14	9	691.0	897	4	0.16
<i>Salvia mellifera</i>	13	390.0	793	5	0.10	17	393.8	789	4	0.12



**Fig. 7.** Spatial pattern of predicted and observed species presence/absence for *Arctostaphylos glauca*, for sample plot locations. Top row: species presence predicted from the CT model, observed, and predicted from the GLM; bottom row: species absence, as above.

function of the predicted probability of species presence,  $p(P)$ , the overall prediction error at the 'optimum' threshold value of  $p(P)$ , when omission and commission errors are both minimized, is always higher than those reported above which are for  $p(P) = 0.5$  (compare Fig. 6 with Tables 5, 6). CT-models based on five variables had overall errors of 5 - 31 % at the optimum value of  $p(P)$ , 0-15% higher than for  $p(P) = 0.5$ , and GLMs had errors of 12-39%, 0-18% higher. Further, the optimum value of  $p(P)$  was roughly equal to how frequently the species occurred in the sample used to develop the model. For example, *Xylococcus bicolor*, *Ceanothus tomentosus*, *Quercus berberidifolia* and *Adenostoma fasciculatum* occurred in 8%, 11%, 46% and 95% of the sample and the optimum  $p(P)$  was 0.1, 0.3, 0.4 and 0.7, respectively.

A simple way to view the spatial distribution of model predictions, prior to mapping them in a GIS

(geographic information system), is to plot the locations of observations from the sample that were predicted present and absent, and compare this to the actual presence/absence of the species in the sample (Fig. 7). In the example shown, the predictions from the classification tree based only on the bioclimatic variables for *Arctostaphylos glauca* are compared to the predictions from the GLM, using the optimum value of  $p(P)$ . The CT-model is more accurate (20% vs. 26% total error); however it shows a pattern of omission errors in the southeast corner of the study area, while the GLM shows even more marked omission errors in the north-west corner (Fig. 7). This is in spite of the fact that the GLM included both spatial variables (utm.n explained the most deviance and utm.e was the fourth-most important), while the CT-model shown here did not include the spatial variables. The information could be used to help identify important additional explanatory variables.

## Discussion

GAMs were useful for examining the shape of the response functions and evaluating them for ecological rationality. These models explained slightly more deviance than GLMs when all variables were included and smoothed. GLMs provided a useful modeling framework for testing alternative hypotheses about the shape of response functions and significance of variables describing environmental gradients. As has been previously noted (McCullough 1995; Austin et al. 1995), even simple tree models (4 - 7 variables, 10 - 20 terminal nodes) can be difficult to interpret in terms of the ecological response functions they approximate and hierarchical interactions they describe.

Many of the models presented here appear to be sufficiently accurate to be used for prediction of species distributions – interpolation within the study area – although they should be evaluated using independent data. All three modeling frameworks, because they predict a probability of class membership – species presence,  $p(P)$  – can be used to map probability surfaces which can be interpreted as an index of habitat suitability (Aspinall & Veitch 1993), or probability of species establishment or occurrence at the site. Presence is more informative than absence in a sample of plant locations (because plants do not occupy every site that they potentially could). However, it is unwise to choose a  $p(P)$  value that minimizes errors of omission, and assume that omission errors simply indicate additional sites where the species could exist but is absent in the sample. This would result in a model that predicts an abundant species to occur everywhere and a rare species almost nowhere (Fig. 6).

Mean minimum temperature of the coldest month was generally the most important climatic variable, followed by mean maximum temperature of the warmest month, and/or annual precipitation. While there is precedent for using these climate variables directly in species response models (Westman 1991; Mackey 1993 for example), bioclimatic indices with a physiological basis may yield results that are more interpretable or comparable to other studies. These include moisture balance (based on precipitation and potential evapotranspiration; cf. Leathwick 1998), growing season warmth (growing degree days) and absolute minimum temperature (Box 1996). This will be addressed in future research.

Previous work suggests that topographic and topohydrologic variables related to soil moisture and nutrient status (slope, hillslope position, upslope catchment area, topographic moisture index, topographically-modeled soil moisture) are also related to plant species distributions and useful in predictive models (reviewed

in Franklin 1995). Only slope was explored in this study, and its poor predictive power may be related to its inaccurate estimation from the USGS DEMs, discussed above. In a previous study using some of the same DEMs but smoothing them using a linear low-pass filter, slope, slope curvature and the topographic moisture index were found to be significant predictors of chaparral species assemblages (McCullough 1995; Franklin et al. in press). Future research will address the accuracy and resolution requirements for digital terrain data used to estimate the variables cited above, in order to thoroughly test their significance in predictive models for these species.

Species response functions can be evaluated not just by their fit, but by their ecological rationality (Austin et al. 1995, p. 77). Climatic variables were frequently modeled using a quadratic response function (see also Westman 1991), although a linear response was not surprising for species sampled over a subset of their biogeographical range. Potential solar insolation was modeled using a linear response function, or, in the case of CT-models, was found at lower levels in the tree, showing a nested hierarchical relationship with the climatic variables (see Michaelsen et al. 1994 for discussion). Spatial variables (geographic position) often showed complex (third and fourth order polynomial) response functions, which may reflect the pattern of the subjectively located plots as much as the spatial distribution of the species, or any biotic clustering or patchiness (Bocard et al. 1992). It may also be related to the spatial patterning in the explanatory variables (Fig. 2) although this remains to be tested. Because of strong clustering of some species among the sample plots, spatial variables accounted for large proportions of the explained deviance in some GAMs and GLMs. However, CT-models that included spatial variables only accounted for a small increase in prediction accuracy over those that did not. While it was noted previously that these models should be used primarily for interpolation within the study area, use of geographic coordinates as explanatory variables would preclude the use of the models for any extrapolation beyond the ecoregion.

**Acknowledgements.** I am indebted to Dena Simons for his help in preparing the data set, Tom White, Hazel Gordon, John Stephenson, Ralph Warbington, Joel Michaelsen, and Ralph Dubayah for providing data, software, guidance and comments, and M.P. Austin, John R. Leathwick, Brendan Mackey, and David Roberts for advice, encouragement, and comments on the manuscript. I also thank the anonymous reviewers and Eddy van der Maarel.

## References

- Aspinall, R. & Veitch, N. 1993. Habitat mapping from satellite imagery and wildlife survey data using a Bayesian modeling procedure in GIS. *Photogramm. Eng. Remote Sens.* 59: 537-543.
- Austin, M.P. 1985. Continuum concept, ordination methods, and niche theory. *Annu. Rev. Ecol. Syst.* 16: 39-61.
- Austin, M.P. 1992. Modelling the environmental niche of plants: some implications for plant community response to elevation CO<sub>2</sub> levels. *Aust. J. Bot.* 40: 615-630.
- Austin, M.P. & Heyligers, P.C. 1989. Vegetation survey design for conservation: gradsect sampling of forests in North-eastern New South Wales. *Biol. Conserv.* 50: 13-32.
- Austin, M.P. & McKenzie, N.J. 1988. Data analysis. In: Gunn, R.H., Beattie, J.A. & Reid, R.E. (eds.) *Australian soil and land survey handbook: Guidelines for conducting surveys*, pp. 210-232. Inkata Press, Melbourne.
- Austin, M.P. & Meyers, J.A. 1995. *Modelling of landscape patterns and processes using biological data, Subproject 4: Real data case study*. Consultancy report to ERIN, CSIRO, Division of Wildlife and Ecology, Canberra.
- Austin, M.P. & Nicholls, A.O. 1997. To fix or not to fix the species limits, that is the ecological question: response to Jari Oksanen. *J. Veg. Sci.* 8: 743-748.
- Austin, M.P. & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35-47.
- Austin, M.P., Meyers, J.A. & Doherty, M.D. 1994a. *Predictive models for landscape patterns and processes, Subproject 2, Modelling of landscape patterns and processes using biological data*. Division of Wildlife and Ecology, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra.
- Austin, M.P., Meyers, J.A., Belbin, L. & Doherty, M.D. 1995. *Modelling of landscape patterns and processes using biological data: Sub-project 5, Simulated data case study*. Division of Wildlife and Ecology, Commonwealth Scientific and Industrial Research Organization, Canberra.
- Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994b. Determining species response functions to an environmental gradient by means of a beta-function. *J. Veg. Sci.* 5: 215-228.
- Austin, M.P., Nicholls, A.O. & Margules, C.R. 1990. Determining the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecol. Monogr.* 60: 161-177.
- Bailey, T.C. & Gatrell, A.C. 1995. *Interactive spatial data analysis*. Longman Scientific & Technical, Harlow, Essex.
- Bayes, A.J. & Mackey, B.G. 1991. Algorithms for monotonic functions and their application to ecological studies in vegetation science. *Ecol. Model.* 56: 135-159.
- Beerling, D.J., Huntley, B. & Bailey, J.P. 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *J. Veg. Sci.* 6: 269-282.
- Bio, A.M.F., Alkemade, R. & Barendregt, A. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *J. Veg. Sci.* 9: 5-16.
- Bocard, D., Legendre, P. & Drapeau, P. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73: 1045-1055.
- Box, E.O. 1996. Plant functional types and climate change at the global scale. *J. Veg. Sci.* 7: 309-320.
- Box, E.O. 1981. *Macroclimate and plant form*. Junk, The Hague.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C. 1984. *Classification and regression trees*. Wadsworth, Belmont, CA.
- Brown, D.B. & Bara, T.J. 1994. Recognition and reduction of systematic error in elevation and derivative surfaces from 7 1/2-minute DEMs. *Photogramm. Eng. Remote Sens.* 60: 189-194.
- Brzeziecki, B., Kienast, F. & Wildi, O. 1995. Modeling potential impacts of climate change on the spatial distribution of zonal forest communities in Switzerland. *J. Veg. Sci.* 6: 257-268.
- Cherrill, A.J., McClean, C., Watson, P., Tucker, K., Rushton, S.P. & Sanderson, R. 1995. Predicting the distributions of plant species at the regional scale: a hierarchical matrix model. *Landscape Ecol.* 10: 197-207.
- Clark, L.A. & Pregibon, D. 1992. Tree-based models. In: Chambers, J. & Hastie, T.J. (eds.) *Statistical models in S*, pp. 377-419. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA.
- Crawley, M.J. 1993. *GLIM for ecologists*. Blackwell Scientific Publications, Oxford.
- Cressie, N. 1991. *Statistics for spatial data*. John Wiley and Sons, New York, NY.
- Daly, C., Neilson, R.P. & Phillips, D.L. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Appl. Meteorol.* 33: 140-158.
- Davis, F.W. & Goetz, S. 1990. Modeling vegetation pattern using digital terrain data. *Landscape Ecol.* 4: 69-80.
- Davis, F.W., Stine, P.A. & Stoms, D.M. 1994. Distribution and conservation status of coastal sage scrub in southwestern California. *J. Veg. Sci.* 5: 743-756.
- Diaz, S. & Cabido, M. 1997. Plant functional types and ecosystem function in relation to global change. *J. Veg. Sci.* 8: 463-474.
- Dozier, J. 1980. A clear-sky spectral solar radiation model for snow-covered mountainous terrain. *Water Resour. Res.* 16: 709-718.
- Dozier, J. 1989. Spectral signature of alpine snow cover from Landsat Thematic Mapper. *Remote Sens. Environ.* 28: 9-22.
- Dozier, J. & Frew, J. 1990. Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Trans. Geosci. Remote Sens.* 28: 963-969.
- Dubayah, R. 1992. Estimating net solar radiation using Landsat Thematic Mapper and digital elevation data. *Water Resour. Res.* 28: 2469-2484.
- Dubayah, R.C. 1994. Modeling a solar radiation topoclimatology for the Rio Grande River Basin. *J. Veg. Sci.* 5: 627-640.
- Dubayah, R. & Rich, P.M. 1995. Topographic solar radiation for GIS. *Int. J. Geogr. Inf. Syst.* 9: 405-419.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474-

- 499.
- Franklin, J. 1996. *Forest Service Southern California Mapping Project: Final Report for the Four Southern Forests*, USFS Contract No. 53-91S8-3-TM45, Unpubl. technical report.
- Franklin, J. & Woodcock, C.E. 1997. Multiscale vegetation data for the mountains of Southern California: spatial and categorical resolution. In: Quattrochi, D.A. & Goodchild, M.F. (eds.) *Scale in remote sensing and GIS*. pp. 141-168. CRC/Lewis Publishers Inc., Boca Raton, FL.
- Franklin, J., McCullough, P. & Gray, C. In press. Terrain variables used for predictive mapping of vegetation communities in Southern California. In: Gallant, J. & Wilson, J. (eds.) *Terrain analysis: principles and applications*. John Wiley & Sons, New York, NY.
- Frew, J.E. 1990. *The image processing workbench*. Ph.D. Thesis, University of California, Santa Barbara, CA.
- Gordon, H. & White, T.C. 1994. *Ecological guide to southern California chaparral plant series*. USDA Forest Service, Pacific Southwest Region, San Diego, CA.
- Guisan, A., Theurillat, J.-P. & Kienast, F. 1998. Predicting the potential distribution of plant species in an alpine environment. *J. Veg. Sci.* 9: 65-74.
- Hanes, T.J. 1977. Chaparral. In: Barbour, M.G. & Major, J. (eds.) *Terrestrial vegetation of California*, pp. 417-469. John Wiley & Sons, New York, NY.
- Harvey, L.E. 1996. Macroecological studies of species composition, habitat and biodiversity using GIS and canonical correspondence analysis. In: *Proceedings, Third International Conference on Integrating GIS and Environmental Modeling*, Santa Fe, NM, January 21-26, 1996, <http://www.ncgia.ucsb.edu/conf/SANTAFE/CD-ROM/main.html>. National Center for Geographic Information and Analysis, Santa Barbara, CA.
- Hastie, T.J. & Tibshirani, R.J. 1990. *Generalized additive models*. Chapman and Hall, London.
- He, H.S. & Mladenoff, D.J. In press. Dynamics of fire disturbance and succession on a heterogeneous forest landscape: a spatially explicit stochastic simulation model. *Ecology*.
- Hickman, J.D. (ed.) 1993. *The Jepson manual: higher plants of California*. University of California Press, Berkeley, CA.
- Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. *J. Biogeogr.* 22: 967-1001.
- Hutchinson, M.F. 1984. *A summary of some surface fitting and contouring programs for noisy data*. CSIRO Division of Mathematics and Statistics and Division of Water and Land Resources, Canberra.
- Hutchinson, M.F. 1987. Methods for generation of weather sequences. In: Bunting, A.H. (ed.). *Agricultural environments: characterisation, classification and mapping*, pp. 149-157. CAB International, Wallingford.
- Hutchinson, M.F., Nix, H.A., McMahon, J.P. & Ord, K.D. 1996. The development of a topographic and climatic database for Africa. In: *Proceedings, Third International Conference on Integrating GIS and Environmental Modeling*, Santa Fe, NM, January 21-16, 1996, <http://www.ncgia.ucsb.edu/conf/SANTAFE/CD-ROM/main.html>. National Center for Geographic Information and Analysis, Santa Barbara, CA.
- Judge, G.G., Griffiths, W.E., Carter Hill, R., Lutkepohl, H. & Tsoung-Chao, L. 1985. *The theory and practice of econometrics*, 2nd ed., John Wiley & Sons, New York, NY.
- Keeley, J.E. & Keeley, S.C. 1988. Chaparral. In: Barbour, M.G. & Billings, W.D. (eds.) *North American terrestrial vegetation*, pp. 165-207. Cambridge University Press, Cambridge.
- Kienast, F., Brzeziecki, B. & Wildi, O. 1996. Long-term adaptation potential of Central European mountain forests to climate change: a GIS-assisted sensitivity assessment. *For. Ecol. Manage.* 80: 133-153.
- Kirkpatrick, J.B. & Hutchinson, C.F. 1980. The environmental relationships of California coastal sage scrub and some of its component communities. *J. Biogeogr.* 7: 23-28.
- Leathwick, J.R. 1995. Climatic relationships of some New Zealand forest tree species. *J. Veg. Sci.* 6: 237-248.
- Leathwick, J.R. 1998. Are New Zealand's *Nothofagus* species in equilibrium with their environment? *J. Veg. Sci.* 9: 721-734. (This issue.)
- Le Duc, M.G., Hill, M.O. & Sparks, T.H. 1992. A method for predicting the probability of species occurrence using data from systematic surveys. *Watsonia* 19: 97-105.
- Lees, B.G. & Ritman, K. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environ. Manage.* 15: 823-831.
- Mackey, B.G. 1993. Predicting the potential distribution of rain-forest structural characteristics. *J. Veg. Sci.* 4: 43-54.
- Mackey, B.G. 1994. A spatial analysis of the environmental relations of rainforest structural types. *J. Biogeogr.* 20: 303-336.
- Mackey, B.G. & Sims, R.A. 1993. A climatic analysis of selected boreal tree species, and potential responses to global climate change. *World Resour. Rev.* 5: 469-487.
- McCullagh, P. & Nelder, J.A. 1989. *Generalized linear models*. Chapman and Hall, London.
- McCullough, P.E. 1995. *Predictive vegetation mapping using classification tree analysis*. M.A. Thesis, San Diego State University, San Diego, CA.
- Michaelsen, J., Schimel, D., Friedl, M., Davis, F.W. & Dubayah, R.C. 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. *J. Veg. Sci.* 5: 673-686.
- Miller, P.C., Hajek, E. & Miller, P.M. 1983. The influence of annual precipitation, topography and vegetative cover on soil moisture and summer drought in southern California. *Oecologia* 56: 385-391.
- Minchin, P.R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69: 89-107.
- Mooney, H.A. 1977. Southern coastal scrub. In: Barbour, M.G. & Major, J. (eds.) *Terrestrial vegetation of California*, pp. 471-489. John Wiley & Sons, New York, NY.
- Moore, I.D., Grayson, R.B. & Ladson, A.R. 1991. Digital terrain modeling: a review of hydrological, geomorphologic

- and biological applications. *Hydrol. Processes* 5: 3-30.
- Nicholls, A.O. 1989. How to make biological surveys go further with generalised linear models. *Biol. Conserv.* 50: 51-75.
- Oksanen, J. 1997. Why the beta-function cannot be used to estimate skewness of species responses. *J. Veg. Sci.* 8: 147-152.
- Pereira, J.M. & Itami, R.M. 1991. GIS-based habitat modeling using logistic multiple regression: A study of the Mt. Graham Red Squirrel. *Photogramm. Eng. Remote Sens.* 57: 1475-1486.
- Pielou, E.C. 1977. *Mathematical ecology*. John Wiley and Sons, New York, NY.
- Poole, D.K. & Miller, P.C. 1981. The distribution of plant water stress and vegetation characteristics in southern California chaparral. *Am. Midl. Nat.* 105: 32-43.
- Ripley, B.D. 1981. *Spatial statistics*. John Wiley and Sons, New York, NY.
- Skidmore, A.K. 1989. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. *Int. J. Geogr. Inform. Syst.* 3: 323-334.
- Smith, P.A. 1994. Autocorrelation in logistic regression modelling of species' distributions. *Global Ecol. Biogeogr. Lett.* 4: 47-61.
- Steffen, W.L., Cramer, W., Plochl, M. & Bugmann, H. 1996. Global vegetation models: incorporating transient changes to structure and composition. *J. Veg. Sci.* 7: 321-328.
- Venables, W.M. & Ripley, B.D. 1994. *Modern applied statistics with S-Plus*. Springer-Verlag, New York, NY.
- Weibel, R. & Heller, M. 1991. Digital terrain modeling. In: Maguire, D.J., Goodchild, M.F. & Rhind, D.W. (eds.) *Geographic Information Systems*, pp. 269-297. Longman, Harlow, Essex.
- Wells, P.V. 1962. Vegetation in relation to geological substratum and fire in the San Luis Obispo Quadrangle, California. *Ecol. Monogr.* 32: 79-103.
- Westman, W.E. 1981. Factors influencing the distribution of species of California coastal sage scrub. *Ecology* 62: 439-455.
- Westman, W.E. 1991. Measuring realized niche spaces: climatic response of chaparral and coastal sage scrub. *Ecology* 72: 1678-1684.
- Woodward, F.I. 1987. *Climate and plant distribution*. Cambridge University Press, Cambridge.
- Woodward, F.I. & Williams, B.G. 1987. Climate and plant distribution at global and local scales. *Vegetatio* 69: 189-197.
- Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* 2: 587-602.
- Zevenberger, L.W. & Thorne, C.R. 1987. Quantitative analysis of land surface topography. *Earth Surface Process. Landform.* 12: 47-56.

Received 18 February 1997;  
 Revision received 12 April 1998;  
 Final revision received 8 July 1998;  
 Accepted 10 August 1998.