Landscape Ecology vol. 4 no. 1 pp 69-80 (1990)
SPB Academic Publishing bv, The Hague

Modeling vegetation pattern using digital terrain data

Frank W. Davis[1] and Scott Goetz[2]
[1] *Department of Geography, University of California, Santa Barbara, CA 93106;* [2] *Science Applications Research, Inc. Greenbelt, MD 20771*

Keywords: California, geographic information system, oak forest, patch size, predictive mapping, remote sensing

Abstract

Using a geographic information system (GIS), digital maps of environmental variables including geology, topography and calculated clear-sky solar radiation, were weighted and overlaid to predict the distribution of coast live oak (*Quercus agrifolia*) forest in a 72 km^2 region near Lompoc, California. The predicted distribution of oak forest was overlaid on a map of actual oak forest distribution produced from remotely sensed data, and residuals were analyzed to distinguish prediction errors due to alteration of the vegetation cover from those due to defects of the statistical predictive model and due to cartographic errors.

Vegetation pattern in the study area was associated most strongly with geologic substrate. Vegetation pattern was also significantly associated with slope, exposure and calculated monthly solar radiation. The proportion of observed oak forest occurring on predicted oak forest sites was 40% overall, but varied substantially between substrates and also depended strongly on forest patch size, with a much higher rate of success for larger forest patches. Only 21% of predicted oak forest sites supported oak forest, and proportions of observed vegetation on predicted oak forest sites varied significantly between substrates. The non-random patterns of disagreement between maps of predicted and observed forest indicated additional variables that could be included to improve the predictive model, as well as the possible magnitude of forest loss due to disturbances in different parts of the landscape.

Introduction

Regional vegetation analyses are conducted routinely by landscape ecologists, geographers and resource managers in order to describe the distribution of plant species and to relate observed distribution patterns to biotic and abiotic site factors (Causton 1988). Typically, vegetation and site measurements from scattered samples are analyzed to develop empirical equations relating vegetation composition to measured site variables. Even in relatively undisturbed areas, such equations or vegetation 'site models' meet with mixed success in predicting actual vegetation patterns because of the complexity and dynamic behavior of plant communities across a range of spatial and temporal scales (Rowe and Sheard 1981). Ground samples inevitably comprise a very small fraction of the mapped region, raising the question of how representative resulting models are for unsampled areas. Samples are of predetermined area deemed suitable for describing vegetation stands, fixing somewhat arbitrarily the spatial scale of the analysis (Noy-Meir and Anderson 1971). Also, samples are usually located subjectively in homogeneous stands selected to be representative of idealized types (*e.g.*, associations, wildlife habitat types, etc.), leading to selective sampling of only some components of actual vegetation cover. As a result, a site model may predict a vegetation pattern very different from the actual pattern over the study region. These predictive errors may have practical consequences when site models are used to project the historical extent of vegetation types, for example to locate restoration projects or natural preserves.

A site model can be tested through additional field sampling; however, there are limits to the amount of field data that can be collected. When maps of site model variables (*e.g.*, geology, topography and soils) exist, a predictive vegetation map can be produced by map weighting and overlaying using a Geographic Information System[1] (GIS). Given a map of actual vegetation distribution, one can overlay the maps to compare predicted to observed vegetation patterns to analyze spatial patterns of disagreement (cf. Thomas 1960).

A number of studies have used GIS capabilities of map weighting and overlay for modeling vegetation pattern based on mapped environmental variables (*e.g.*, Box 1981). Most recently, predictive vegetation maps have been used in remote sensing applications to improve land cover classifications based on digital satellite data (*e.g.*, Strahler 1981; Morissey and Strong 1986; Cibula and Niquist 1987). In these studies, predictive models were developed from ground samples and the GIS was used to extrapolate across unsampled areas. Our research approach is similar, except that we are concerned with comparing predicted vegetation patterns to independently derived maps of actual vegetation (*e.g.*, Hill and Kelly 1987).

In principle, the interpretation of residual patterns from a comparison of observed and predicted vegetation maps is extremely complicated, because predictive errors can originate both from errors in maps of site variables and actual vegetation, and from inadequacies of the site model. We have found in practice, however, that residual patterns may be interpretable based on the analysts' knowledge of the data sources and the region under investigation, supplying much information not obtainable from simple goodness-of-fit statistics or additional field sampling. For example, patterns in residuals may reveal model biases, ecological subregions or ecological variables not previously recognized. Furthermore, knowing how a vegetation model performs in different parts of the study region can temper its application to management and planning decisions.

We have used digital maps of site variables (*i.e.*, geology, topography and calculated clear-sky solar radiation) and GIS capabilities to map the predicted distribution of natural vegetation types in coastal California whose actual distributions were mapped using Thematic Mapper Simulator (TMS) data. We compared the actual distribution of one vegetation type, coast live oak (*Quercus agrifolia* Neé) forest, to the distribution predicted by a quantitative site model, to answer the following questions:

- What is the total area and patch size distribution of observed oak forest?
- What is the total area and patch size distribution of predicted oak forest?
- For areas of observed oak forest, what is the amount and patch size distribution of predicted vegetation types?
- For areas of predicted oak forest, what is the amount and patch size distribution of observed vegetation types?
- How are areas where predicted and observed maps disagree distributed with respect to geology and topography?

Our overriding objectives in this paper are to test the power of mapped site variables for predicting the distribution of natural vegetation in coastal California, to demonstrate the utility of high resolution satellite data and GIS capabilities in regional vegetation analyses, and to call attention to some methodological issues of data scale and data quality that must be addressed in applying these technologies to regional vegetation modeling.

[1] Burrough (1986, p. 6) defines a GIS as 'a set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world for a particular set of purposes.'

Study area

We modeled natural vegetation pattern over a 72 km^2 area northeast of Lompoc, California (latitude 34°42′N, longitude 120°27′W). The climate here is mediterranean, with relatively cool summers and mild winters. Over 90% of the 36 cm average annual precipitation falls between November and April.

Two distinct physiographic regions occur in the study area; Burton Mesa and the Purisima Hills. Burton Mesa is a marine terrace underlain by marine sedimentary rocks that are covered with Orcutt sandstone, 0.5–40 meters of weakly cemented Quaternary aeolian sands (Diblee 1950). Level upland expanses from 100–120 m above sea level are separated by wide valleys filled with Quaternary alluvium.

Most vegetated areas are covered by maritime chaparral, which is dominated by evergreen shrub species including *Adenostoma fasciculatum, Ceanothus ramulosus, Arctostaphylos rudis* and *A. purisima* (Davis *et al.* 1988). Multi-stemmed coast live oaks 3–6 m in height are interspersed throughout the chaparral, attaining 40–70% crown cover in areas not recently disturbed by burning or clearing. Coastal sage scrub and annual grassland occur on formerly cleared sites and on south-facing slopes. Coast live oak forest is most extensive on steep north-facing slopes and in riparian corridors.

The Purisima Hills are a northwest-southeast trending anticline of marine sedimentary rocks. Elevations range from 225 to 450 m, and topography consists of rolling hills with short steep slopes. Important geologic formations in the study area include the Sisquoc diatomite and shale, the Careaga sandstone and the Paso Robles conglomerate. Predominant vegetation types in the Purisima Hills include coastal sage scrub, chaparral, bishop pine (*Pinus muricata*) forest, coast live oak woodland and coast live oak forest. Vegetation pattern is associated strongly with geology and topography. Cole (1980) documented the association of bishop pine forest with the diatomaceous member of the Sisquoc Formation, coast live oak forest with north facing slopes of the Careaga sandstone and Sisquoc shale, and coastal sage scrub or chaparral with steep south facing slopes of the Purisima Hills.

Natural vegetation in the study area is fragmented by roads, residential areas, agriculture and other developments. Remaining vegetation has experienced a complex disturbance history over the past century or more that includes wildfire, grazing and clearing. These disturbances exert a strong and persistent effect on vegetation composition and weaken the association between actual vegetation and mapped site variables (*e.g.*, Wells 1962; Davis *et al.* 1988). We applied predictive mapping only within areas where actual vegetation was dominated by native shrub or tree species. We excluded annual grasslands, nearly all of which were either actively grazed or recently burned or cultivated (see below).

Although we modeled the distribution of 5 vegetation types (Table 1), we focused on the actual and predicted distribution of coast live oak (*Quercus agrifolia* Neé) forest, which we define as vegetation where the species attains at least 60% canopy cover. Because coast live oak is the only dominant broadleaf evergreen tree in the study area, vegetation containing the species has a distinctive reflectance and can be mapped reliably with high resolution satellite data and aerial photography (Davis 1987). Furthermore, because coast live oak is relatively less adapted to drought than other mediterranean plant species, oak forests are generally restricted to mesic substrates and sites such as steep north-facing slopes and riparian corridors (Wells 1962; Griffin 1973; Cole 1980). The documented association of the species with mapped surficial geology and topography makes it especially suited for testing the potential of GIS-based predictive mapping.

Methods

A vegetation map for the study area was produced using Thematic Mapper simulator data (28 m resampled to 30 m resolution) collected in July 1984 (Davis 1987). Natural vegetation classes were mapped with 89% accuracy overall (accuracy determined following Card (1982); see Davis (1987) for details). All classes were mapped with greater than 85% accuracy except for oak forest, which was

Table 1. Classification system used to map dominant natural vegetation types in the study region. For logistic regression analysis, oak woodland and chaparral were merged into a 'woodland/chaparral' category. Grassland and Willow woodland were excluded from the analysis. Map accuracy for each class is the proportion of samples classified correctly in the TMS-derived vegetation map, based on 141 test sites (see Davis (1987) for details).

Class	% Oak cover	Dominant species	Map accuracy (%)
Coast live oak forest	>60	*Quercus agrifolia* *Toxicodendron diversilobum*	79
Coast live oak woodland	20–60	*Quercus agrifolia* *Adenostoma fasciculatum* *Arctostaphylos spp.*	86
Chaparral	0–20	*Quercus agrifolia* *Adenostoma fasciculatum* *Ceanothus ramulosus* *C. impressus* *Arctostaphylos rudis* *A. purisima*	89
Coastal Scrub	*0–20*	*Salvia mellifera* *Baccharis pilularis* *Ericameria ericoides* *Artemisia californica*	86
Conifer Forest	0–30	*Pinus muricata* *Quercus agrifolia* *Heteromeles arbutifolia*	92
Grassland	0–20	*Bromus spp.* *Vulpia spp.* *Avena barbata* *Brassica spp.*	89
Willow woodland	0–20	*Salix spp.*	100

mapped with 79% accuracy (Table 1). Oak forest was most frequently confused with dense oak woodland. This is not a severe mapping error, given that one class grades into the other.

The vegetation map was co-registered in Universal Transverse Mercator (UTM) projection to a geologic map of the study area (Dibblee 1950) that we digitized using Earth Resources Data Analysis System (ERDAS) software (Fig. 1). Dibblee originally mapped 19 geologic series at 1:50,000 scale. We did not attempt to quantify the accuracy of the map. To simplify the analysis of association between vegetation and geology, recent Quaternary deposits, including terraces, alluvium and Orcutt sandstone, were merged into a single class (Orcutt sand comprised 86% of this class). All three series were characterized by deep sandy soils. We analyzed three other widespread lithologic units, including the Paso Robles conglomerate, Careaga sandstone, and Sisquoc diatomaceous shale. Although soil maps exist for the study area, we did not use them because the soil maps for the Purisima Hills were less detailed than the geologic map and had less predictive value.

Topographic variables including elevation, slope angle and slope aspect, clear-sky solar radiation and drainage area were derived from the U.S. Geological Survey 30 m digital elevation model (DEM) for the Lompoc quadrangle using software developed at the UCSB Department of Geography (Frew and Dozier 1986). Unsmoothed elevations possessed 1 m vertical and 30 m horizontal resolution, with a nominal root mean square error of 3.0 m in both vertical and horizontal dimensions. Based on transit surveys of several hillslope profiles on eastern Burton Mesa, there was good agreement between actual and mapped elevations ($r^2 = 0.93$), but only fair agreement between actual and mapped

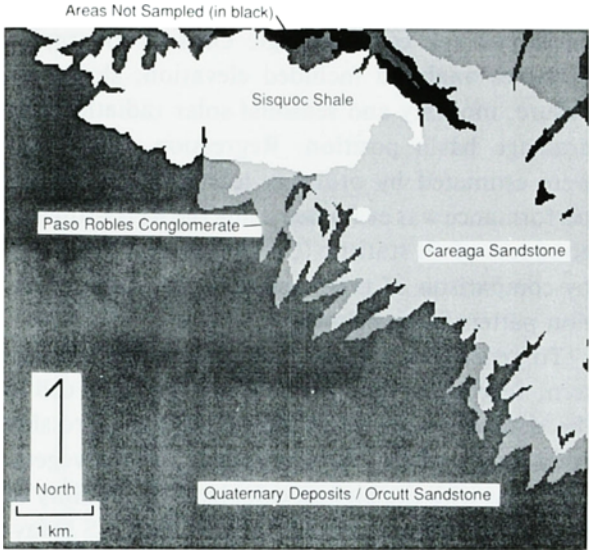

Fig. 1. Surficial geology of the study area (simplified from Dibblee 1950).

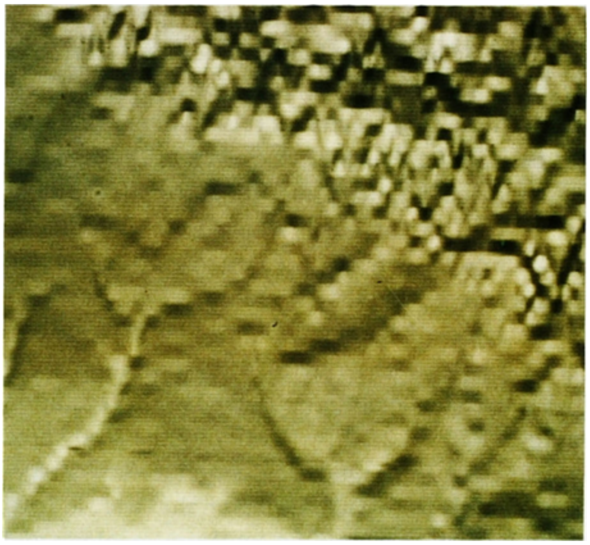

Fig. 2. Distribution of integrated January insolation calculated from digital elevation data. Image brightness is proportional to total insolation. Image orientation and area are the same as in Fig. 1.

slope angle ($r^2 = 0.41$) and slope aspect ($r^2 = 0.38$) (Goetz 1987). This is partly because errors in elevation data were amplified by the local differencing operations used to calculate slopes and exposures. Errors were concentrated in areas of rapidly changing slope and exposure such as ridges and ravines, and included both resolution errors (*i.e.*, undersampling in areas of rapid change) and stereomodel errors (*e.g.*, overestimating surface elevation in riparian corridors filled with tall, continuous tree canopy).

Incident radiation on a slope was calculated using maps of slope angle and slope aspect as well as a horizon file which provided, for each cell in the elevation model, the angle to the local horizon for 8 different azimuth sectors (*i.e.*, north, northeast, . . .) (Dozier 1980; Dozier *et al.* 1981). Terms for diffuse irradiance and reflected radiation from surrounding terrain were estimated under specified conditions of atmospheric scattering and transmittance and surface albedo. The range in elevations was small enough that the atmosphere was treated as the same at all locations.

To produce maps of monthly solar radiation for the months of December through June, we calculated instantaneous radiation at hourly intervals for three days in each month, and integrated these

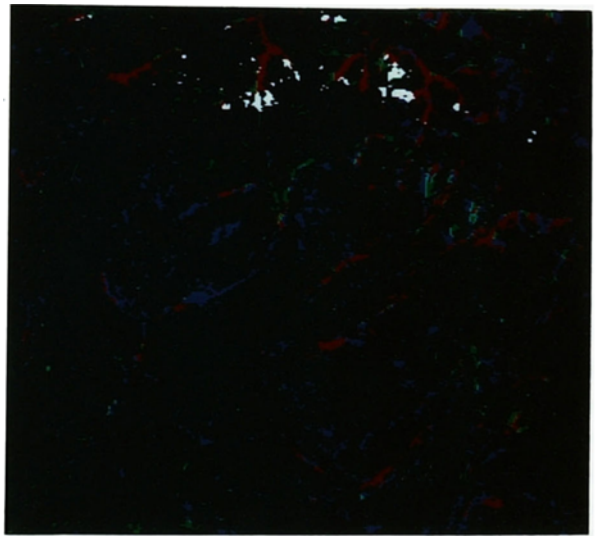

Fig. 3. Distribution of coast live oak forest mapped using July, 1984 TMS data, shaded to indicate predicted vegetation types on observed oak forest. Black areas are non-forested areas. Colored areas are existing oak forest that were predicted by the logit regression model to be oak forest (red), oak woodland (blue), coastal scrub (green) or conifer forest (white). Image orientation and area are as in Fig. 1.

values over the entire month. Because we could only roughly estimate seasonal atmospheric properties, the calculated insolation values were treated as relative and scaled from 0 to 255 (Fig. 2).

Variation in soil moisture related to drainage basin position (*e.g.*, upper slope versus lower slope) was modeled by calculating, for each cell, the number of cells in the basin which were expected to drain through that cell based on maps of slope and exposure (*e.g.*, Band 1986; Marks *et al.* 1984).

The association of vegetation and mapped terrain variables was modeled using polychotomous logit regression analysis (Wrigley 1975). Vegetation samples were located by stratifying the study area into six subregions, and then sampling 40–60 vegetation stands from each subregion that were at least 60 by 60 meters in area on 'uniform' geology and topography (to minimize cartographic error). Sample neighborhood was selected randomly, but sample locations were sometimes adjusted 30–60 m to meet our criteria of uniform vegetation and site conditions. Vegetation type and percent cover by coast live oak in each sample were determined using 1983 1:24,000 color aerial photography (high photointerpretation accuracy was verified during numerous field visits between 1985 and 1987). Geologic substrate and values for topographic variables were taken from the digital database.

The data consisted of 258 samples of four vegetation types, oak forest ($n = 60$), oak woodland and hard chaparral ($n = 116$), coastal scrub ($n = 62$) and conifer forest ($n = 20$). We excluded willow woodland because it is infrequent and is associated with riparian areas that we could not model successfully using the DEM data. As mentioned above, we also excluded grassland because this type occurs nearly exclusively on recently disturbed sites. Initially, oak woodland and chaparral were analyzed separately, but we observed no difference in the site relations of these two types so these types were combined to increase class sample size for estimating logit model coefficients. Oak cover increases during fire-free intervals on many chaparral-covered sites in the study area, and on these sites chaparral is probably seral to woodland (Wells 1962; Davis *et al.* 1988).

Initial data exploration indicated that site relations of the vegetation classes differed among the substrates, so separate logit regression models were developed for each geologic class. Topographic variables analyzed included elevation, slope, exposure, monthly and seasonal solar radiation and drainage basin position. Regression coefficients were estimated by ordinary least squares. Model performance was evaluated using the RHO-squared goodness-of-fit statistic (Costanzo *et al.* 1982) and by comparison of predicted and observed vegetation patterns (see below).

To generate a map of predicted vegetation pattern, vegetation class probabilities for each cell in the database were calculated from the regression equations, and the cell was assigned to the vegetation class with the highest calculated probability of occurrence using the program PROBCLAS (Maynard and Strahler 1981).

The correspondence between maps can be measured by testing for non-random distribution of map residuals using spatial measures of contiguity or spatial autocorrelation (Cliff and Ord 1981), or using aspatial measures of contingency or correlation (Phipps 1981). Given the large sample size ($n = 79{,}605$ cells) we assessed map correspondence using non-spatial analyses of randomly located samples. The use of conventional significance tests of association was problematic because the spatial dependence in mapped variables violated the assumption of sample independence (Fingleton 1986). To avoid this problem we sub-sampled the maps at a sampling density low enough so that sample values were expected to be independent at the average intersample distance. For topographic variables, the sampling distance was determined empirically by semi-variogram analysis (Oliver and Webster 1986) to be around 210 m, corresponding to a 2% sample of the region. Accordingly, the association of observed vegetation pattern with topographic variables was measured for a random sample of 1450 cells (1.8% of the study region) from the database.

Results

Oak forest was mapped over 4.5% of the study area (Fig. 3). The remaining area was mapped as oak woodland and chaparral (19.4%), coastal

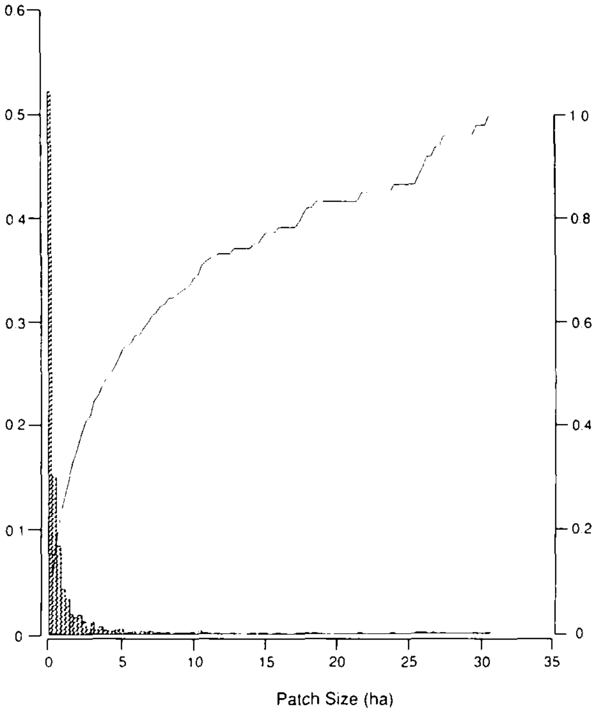

Fig. 4. Patch size distribution of observed oak forest in classified TMS image (bars) and cumulative proportion of forested area as a function of patch size (line).

scrub (20.0%), conifer forest (2.9%), or other (residential, cropland, grassland, willow woodland) (53.2%). Mapped stands of oak forest averaged 0.51 ha, with the size distribution strongly skewed towards the 0.09 ha resolution of the TMS data (Fig. 4). Some of the small patches were local dense clusters of oaks in stands of oak woodland and chaparral (Davis 1987). These occurred primarily on Burton Mesa. Other small patches were forest stands that were highly localized in riparian corridors or mesic coves, or were remnant fragments in areas subjected to historical clearing and burning.

Table 3. Summary of polychotomous logit regression models for four potential natural vegetation classes, Burton Mesa and Purisima Hills. Signs in parentheses indicate the direction of the relationship between the topographic variable and the likelihood of oak forest.

Geologic substrate	Significant variables	RHO-squared
Quarternary deposits	March insolation (−) Slope (−)	0.246
Paso robles conglomerate	March insolation (−)	0.174
Careaga sandstone	March insolation (−) Aspect (+)	0.228
Sisquoc Diatomite	December insolation (−) Elevation (−)	0.192
All substrates		0.338

Overlaying maps of geology and vegetation corroborated the observations by Cole (1980) that conifer forest in the region is essentially restricted to diatomaceous shale of the Sisquoc formation (Table 2). Stratification of the region by geology combined with logit regression models based on topographic variables gave a relatively high RHO-squred of 0.338 (Table 3). The separate logit regression models had only moderate predictive skill, with values for RHO-squared of 0.17–0.25. Calculated March radiation was the topographic variable most strongly associated with the pattern of natural vegetation on all substrates except the Sisquoc diatomite, where December radiation was a better predictor. Differences in the association of vegetation pattern and solar radiation for the months of December through March were slight (correlation of March and December radiation = 0.97).

The RHO-squared statistics indicated how well the model fit the 258 training samples, but a more

Table 2. Frequencies and relative percentages of 4 natural vegetation classes and other land cover types on four geologic substrates in the study area (n = 79,605 cells). Percentages for each substrate sum to 100%.

Geology	Oak forest		Oak woodland/chaparral		Coastal scrub		Conifer forest		Other	
Quaternary deposits	2423	0.05	18824	0.39	8572	0.18	4	0.00	18530	0.38
Paso Robles conglomerate	588	0.12	1335	0.26	1343	0.26	3	0.00	1811	0.36
Careaga sandstone	1927	0.11	6468	0.37	6195	0.36	8	0.00	2848	0.16
Sisquoc shale	939	0.11	5034	0.58	1434	0.16	500	0.06	819	0.09

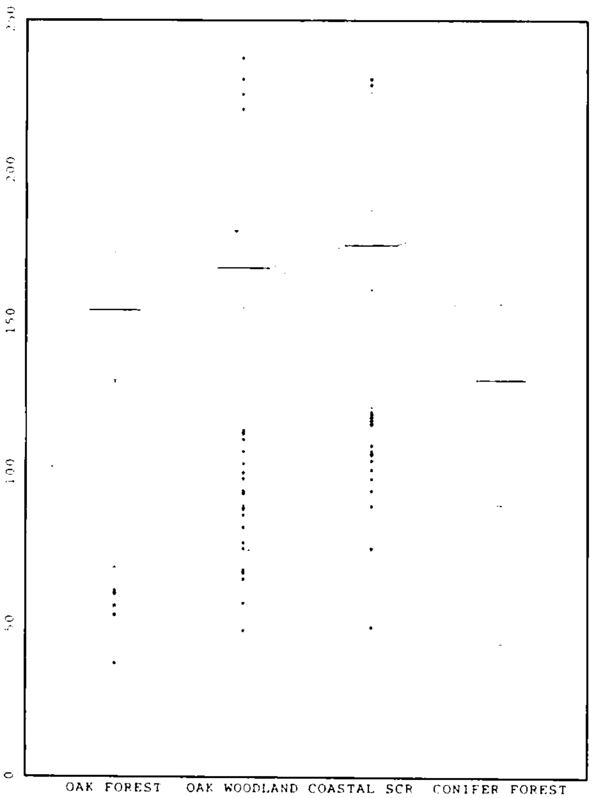

Fig. 5. Boxplots showing the distribution of March insolation for 4 vegetation classes on all geologic substrates, based on a random sample of 1450 cells from the database. Sharp ridges and ravines were excluded from the sample, because of the lower accuracy of DEM data in those areas. Boxes show the upper quartile, median and lower quartile for observations; vertical lines and asterisks show upper and lower extremes and outliers. Non-overlapping of notches indicates difference at a rough 5% significance level (Chambers *et al.* 1983).

general test of model performance was provided by comparing predicted to observed vegetation patterns for the entire study region. The proportion of observed oak forest that occurred on predicted oak forest sites was 40% overall, but varied substantially between substrates (Table 4). For example, most observed oak forest on Quaternary deposits mapped onto predicted oak woodland sites. Low predictive success of model was the result of: 1) cartographic error due to confusion of dense oak woodland and oak forest in the map of actual vegetation, and 2) ecological error, in the sense that oak forest was not as restricted to low radiation environments as the model predicted. For example, many small patches of oak forest were predicted oak woodland sites on level uplands of Burton Mesa that were not recently burned or cleared.

The proportion of observed oak forest on predicted oak forest sites also depended strongly on patch size (Fig. 6), with a much higher rate of success for larger patches of forest. Excluding patches less than 2 hectares (58% of mapped oak forest), 60% of remaing forest occurred on predicted oak forest sites. The three largest patches of oak forest, all greater than 10 ha in size, fell entirely within predicted oak forest areas. Although we could not account fully for this scale-dependence in model fit, it was due in part to the high error rate for small oak forest patches on Quarternary deposits. Also, larger patches of oak forest tended to occur on larger more homogeneous slopes, which were more accurately depicted by the DEM data. Finally, we observed in the field that many smaller patches of mapped oak forest occurred near seeps, along geologic contacts, in swales and near lower order streams, all environments that were not depicted reliably by the database.

Only 21% of predicted oak forest sites supported oak forest. 55% supported oak woodland and chaparral and 24% supported coastal scrub, conifer forest or other cover types (mainly grassland, cropland and residential) (Fig. 7). Proportions of observed vegetation on predicted oak forest sites varied sharply between substrates (Table

Table 4. Relative proportions of observed oak forest on predicted vegetation types as a function of substrate type (columns sum to 1).

Predicted vegetation	Quaternary deposits	Paso Robles conglomerate	Careaga sandstone	Sisquoc shale
Oak forest	0.24	0.27	0.42	0.50
Oak woodland	0.73	0.37	0.42	0.02
Coastal scrub	0.03	0.36	0.16	0.13
Conifer forest	0.00	0.00	0.0	0.35

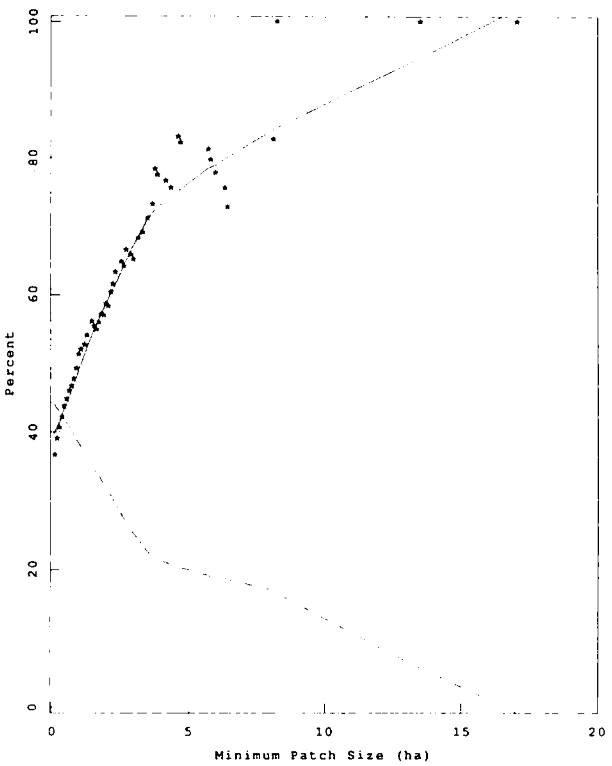

Fig. 6. Percent of observed oak forest occurring on predicted oak forest sites as a function of minimum forest patch size analyzed. Asterisks are actual data values. Solid line was fitted using locally weighted regression (Chambers *et al.* 1983). Broken line shows corresponding percent of observed oak forest on predicted oak woodland sites.

5). For example, most predicted oak forest on Sisquoc shale was observed to be oak woodland and chaparral, whereas on the Paso Robles it was mainly coastal scrub and other. This partly reflected differences in land use and disturbance on these substrates. Fire has been the major form of disturbance on Sisquoc shale, whereas large areas of the Paso Robles conglomerate and Careaga sandstone have been cleared and grazed. On several substrates the residuals were systematically associated with different topographic variables. For example, on the Sisquoc shale, conifer forest and oak woodland/chaparral occurred at significantly higher elevation than oak forest on predicted oak forest sites. We attributed this result to the association of these vegetation types with the diatomaceous member of the Sisquoc Formation, which occupes higher elevations. Thus this model bias could be reduced by including a more detailed geologic classification.

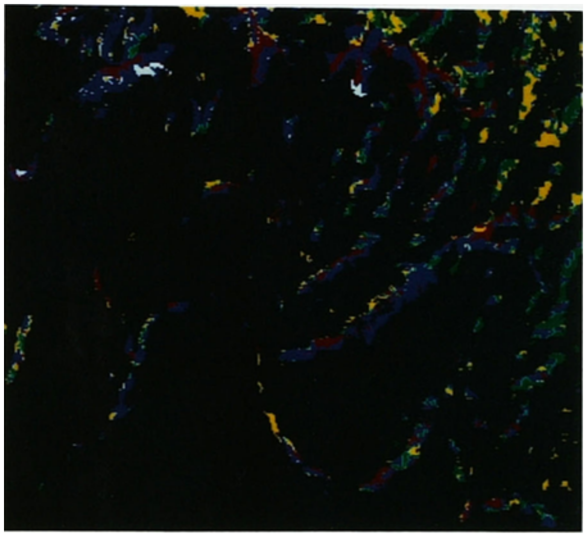

Fig. 7. Predicted distribution of coast live oak forest based on geology, topography and insolation. Black areas are predicted vegetation other than oak forest. Colored areas are predicted oak forest sites on which mapped existing vegetation was oak forest (red), oak woodland (blue), coastal scrub (green), conifer (white) or other land cover types (yellow). Image orientation and area are as in Fig. 1.

Discussion

The association between vegetation and calculated monthly radiation was relatively strong for the months of December through March, in spite of the inaccuracies and relatively coarse resolution of the DEM data. Also, vegetation pattern was more strongly associated with calculated radiation than with measures of slope orientation that did not account for shading by local horizons. These results indicate the potential for analyzing plant species distributions in relation to dynamic patterns of solar radiation using high resolution (*e.g.*, 5–10 m) digital elevation data. Previously such analyses were possible only for sample points or localized transects (*e.g.*, Kirkpatrick and Nunez 1980). Using accurate higher resolution data it should also be possible to relate vegetation patterns to topographically-controlled patterns in soil moisture or surface hydrology (*e.g.*, O'Loughlin 1986; Band 1986).

Results of predictive mapping suggest that coast

Table 5. Relative proportions of observed vegetation or land cover types on areas predicted as oak forest sites, as a function of substrate type (columns sum to 1).

Observed vegetation	Quaternary deposits	Paso Robles conglomerate	Careaga sandstone	Sisquoc shale
Oak forest	0.17	0.14	0.18	0.13
Oak woodland	0.54	0.24	0.35	0.65
Coastal scrub	0.14	0.34	0.33	0.05
Conifer forest	0.00	0.00	0.00	0.04
Other	0.14	0.28	0.15	0.13

live oak forest occupies only a small fraction of existing suitable habitat in the region, and that in most areas it has been replaced by oak woodland and chaparral. Wells (1962) blamed anthropogenic fires, cutting and grazing for the conversion of large areas of oak forest to chaparral in this area. Oak forest may require several to many decades to recover from such disturbances (Davis *et al.* 1988), although the rates probably vary between sites and depending on the nature of the disturbance. In this study we observed that much of the observed oak forest occurred only on the lower portion of slopes that were predicted oak forest. This could be a systematic flaw in the predictive site model, an indication of less frequent or less intense disturbance (especially fire) of lower hillslope areas, or more rapid recovery of oak forest from disturbances in these sites. Including maps of recent fire and land use histories should help to resolve some of the discrepancies between observed and predicted vegetation patterns in the region.

Our analyses of coast live oak forest are based on relatively simple GIS operations combining map weighting and overlay, patch size analysis, and spatial sampling. Such GIS-based ecological analyses are useful to the degree that maps derived from a sequence of cartographic operations are of sufficient spatial resolution to describe the ecological processes under investigation, and are of sufficient accuracy so that ecological information is not overwhelmed by cartographic noise. Digital maps contain inaccuracies due both to inherent errors in the original data and operational errors from map digitizing and registration (Burrough 1986; Walsh *et al.* 1987), so that a geographical database is at best a 'fuzzy' representation of the landscape (Robinson and Strahler 1984). For this reason there is a trade-off between model complexity (*e.g.*, more variables or more complex spatial operations) and model reliability (Burrough 1986). In the analyses reported here, we could readily generate enough samples from the database to outweigh cartographic errors, so that previously documented associations of vegetation pattern with geology and topography were detectable (*e.g.*, Wells 1962; Harrison *et al.* 1971; Cole 1980; Westman 1981).

The results presented above are intended to illustrate how GIS-based cartographic modeling can contribute to the analysis of regional vegetation patterns and the association of vegetation with environmental variables. We are not suggesting that cartographic modeling can substitute for field sampling in developing and testing vegetation site models. However, the types of cartographic analyses conducted here complement traditional field survey methods by measuring associations or testing field results with many more random samples and at larger spatial scales than can practically be collected in the field, facilitating the analysis of large heterogeneous landscapes. Furthermore, we have found that the ability to overlay predicted on observed landscape patterns gives a strong sense of the true predictive skill and bias of quantitative site models, providing useful guidance in terms of model improvement and application.

Acknowledgements

Research support was provided by the California Space Institute. We are grateful to Carol Johnston and two anonymous reviewers for providing helpful comments on the draft manuscript.

References

Band, L.E. 1986. Topographic partition of watersheds with digital elevation models. Wat. Resour. Res. 22: 15–23.

Box, E.O. 1981. Macroclimate and plant forms: an introduction to predictive modeling in phytogeography. Dr W. Junk, The Hague.

Burrough, P.A. 1986. Principles of geographic information systems for land resource assessment. Clarendon Press, Oxford.

Card, D.H. 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogramm. Engng. and Rem. Sens. 48: 431–439.

Causton, D.R. 1988. Introduction to vegetation analysis. Unwin Hyman, London.

Chambers, J.M., Cleveland, W.B., Kleiner, B. and Tukey, P.A. 1983. Graphical methods for data analysis. Duxbury Press, Boston.

Cibula, W.G. and Nyquist, M.O. 1987. Use of topographic and climatological models in a geographical data base to improve Landsat MSS classification for Olympic National Park. Photogramm. Engng. and Rem. Sens. 53: 67–76.

Cliff, A.D. and Ord, J.K. 1981. Spatial processes. Pion, London.

Cole, K. 1980. Geologic control of vegetation in the Purisima Hills, California. Madrono 27: 79–89.

Costanzo, C.M., Halperin, W.C., Gale, N.D. and Richardson, G.D. 1982. An alternative method for assessing goodness-of-fit for logit models. Environ. and Plann. A 14: 963–971.

Davis, F.W. 1987. Thematic mapper analysis of coast live oak in Santa Barbara County. *In*: Proceedings of the Symposium on Multiple-use Management of California Hardwood Resources, pp. 317–324. United States Forest Service Pacific Southwest Range Experiment Station Gen. Tech. Rep. PSW-100. Berkeley, California.

Davis, F.W., Hickson, D.E. and Odion, D.C. 1988. Composition of maritime chaparral related to fire history and soil, Burton Mesa, California. Madrono 35: 169–195.

Dibblee, T.W. 1950. Geology of southwestern Santa Barbara County, California Division of Mines. Bulletin 150. Sacramento, California.

Dozier, J. 1980. A clear-sky spectral solar radiation model for snow-covered mountainous terrain. Water Resour. Res. 16: 709–718.

Dozier, J., Bruno, J. and Downey, P. 1981. A faster solution to the horizon problem. Comput. Geosci. 7: 145–151.

Fingleton, B. 1983. Log-linear models with dependent spatial data. Environ. and Plann.A 15: 801–813.

Frew, J. and Dozier, J. 1986. The image processing workbench – portable software for remote sensing instruction and research. *In*: Proceedings of the 1986 International Geoscience and Remote Sensing Symposium, ESA SP-254. pp. 271–276. European Space Agency, Paris.

Goetz, S. 1987. Predictive mapping of Coast live oak in California using digital terrain data. M.A. Thesis, University of California, Santa Barbara.

Griffin, J.R. 1973. Xylem sap tension in three woodland oaks of central California. Ecology 54: 862–868.

Harrison, A.T., Small, E. and Mooney, H.A. 1971. Drought relationships and distribution of two mediterranean-climate California plant communities. Ecology 52: 869–875.

Hill, G.J.E. and Kelly, G.D.1987. A comparison of existing map products and landsat for land cover mapping. Cartography 16: 51–57.

Kirkpatrick, J.B. and Nunez, M. 1980. Vegetation-radiation relationships in mountainous terrain: eucalypt-dominated vegetation in the Risdon Hills, Tasmania. J. Biogeogr. 7: 197–208.

Marks, D., Dozier, J. and Frew, J. 1984. Automated basin delineation from digital terrain data. Geo-Processing 2: 299–311.

Maynard, P.F. and Strahler, A.H. 1981. The logit classifier: a general maximum likelihood discriminant for remote sensing applications. *In*: 15th International Symposium on Remote Sensing of the Environment. pp. 1009–1026. Ann Arbor, Michigan.

Morissey, L.A. and Strong, L. 1986. Mapping permafrost in the Boreal Forest with thematic mapper satellite data. Photogramm. Engng. and Rem. Sens. 52: 1513–1520.

Noy-Meir, I. and Anderson, D.J. 1971. Multiple pattern analysis, or multiscale ordination: towards a vegetation hologram? *In*: Statistical Ecology III: Many Species Populations, Ecosystems and Systems Analysis. pp. 207–231. Edited by G.P. Patil, E.C. Pielou and W.E. Waters. Pennsylvania State University Press, University Park.

Oliver, M.A. and Webster, R. 1986. Semi-variograms for modelling the spatial pattern of landform and soil properties. Earth Surf. Processes Landforms 11: 491–504.

O'Loughlin, E.M. 1986. Prediction of surface saturation zones in natural catchment by topographic analysis. Wat. Resour. Res. 22: 794–804.

Paysen, T.E., Derby, J.E. , Black, H., Jr., Bleich, C. and Mincks, J.W. 1980. A vegetation classification system applied to Southern California. Gen. Tech. Rep. PSW-45. United States Forest Service Pacific Southwest Range Experiment Station. Berkeley, California.

Phipps, M. 1981. Entropy and community pattern analysis. J. Theor. Biol. 93: 253–273.

Robinson, V.B. and Strahler, A.H. 1984. Issues in designing geographic information systems under conditions of inexactness. *In*: Symposium on Machine Processing of Remotely Sensed Data. pp. 198–204. IEEE, New York.

Rowe, J.S. and Sheard, J.W. 1981. Ecological land classification: a survey approach. Environ. Manage. 5: 451–464.

Strahler, A.H. 1981. Stratification of natural vegetation for forest and rangeland inventory using Landsat digital imagery and collateral data. Int. J. Rem. Sens. 2: 15–41.

Thomas, E.W. 1960. Maps of residuals from regression: their characteristics and uses in geographic research. Department of Geography Publication No. 2, State University of Iowa, Iowa City.

Walsh, S.J., Lightfoot, D.R. and Butler, D.R. 1987. Recognition and assessment of error in geographic information sys-

tems. Photogramm. Engng. and Rem. Sens. 53: 1423–1430.
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. Diversity relations and succession in California coastal sage scrub. Ecology 62: 170–184.
Wrigley, N. 1975. Analyzing multiple alternative dependent variables. Geogr. Anal. 7: 187–195.