



Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA

Emma C. Underwood*, Rob Klinger† and Peggy E. Moore‡

*Department of Environmental Science & Policy, University of California, Davis, California 95616, USA, †Section of Evolution & Ecology, University of California, Davis, California 95616, USA, ‡Western Ecological Research Center, US Geological Survey, PO Box 700, El Portal, California 95318, USA

ABSTRACT

One of the major issues confronting management of parks and reserves is the invasion of non-native plant species. Yosemite National Park is one of the largest and best-known parks in the United States, harbouring significant cultural and ecological resources. Effective management of non-natives would be greatly assisted by information on their potential distribution that can be generated by predictive modelling techniques. Our goal was to identify key environmental factors that were correlated with the percent cover of non-native species and then develop a predictive model using the Genetic Algorithm for Rule-set Production technique. We performed a series of analyses using community-level data on species composition in 236 plots located throughout the park. A total of 41 non-native species were recorded which occurred in 23.7% of the plots. Plots with non-natives occurred most frequently at low- to mid-elevations, in flat areas with other herbaceous species. Based on the community-level results, we selected elevation, slope, and vegetation structure as inputs into the GARP model to predict the environmental niche of non-native species. Verification of results was performed using plot data reserved from the model, which calculated the correct prediction of non-native species occurrence as 76%. The majority of the western, lower-elevation portion of the park was predicted to have relatively low levels of non-native species occurrence, with highest concentrations predicted at the west and south entrances and in the Yosemite Valley. Distribution maps of predicted occurrences will be used by management to: efficiently target monitoring of non-native species, prioritize control efforts according to the likelihood of non-native occurrences, and inform decisions relating to the management of non-native species in postfire environments. Our approach provides a valuable tool for assisting decision makers to better manage non-native species, which can be readily adapted to target non-native species in other locations.

Keywords

Biological invasions, GARP, invasive plants, non-native species, park management, plant community analysis, predictive modelling.

*Correspondence: Emma Underwood, Department of Environmental Science & Policy, One Shields Avenue, University of California, Davis, California 95616–8571, USA.
E-mail: eunderwoodrussell@ucdavis.edu

INTRODUCTION

One of the key challenges confronting ecologists and land managers is how to manage non-native species in order to conserve native biodiversity and maintain traditional ecosystem functions. The effect of invasive non-native species in natural systems is generally regarded as one of the most critical issues confronting conservation science (Drake *et al.*, 1989; Simberloff *et al.*, 1997; Mack *et al.*, 2000). The ecological effects of non-native species have been well documented at the population, community, and ecosystem levels (Vitousek, 1986; Gordon, 1998; Mack & D'Antonio, 1998; Parker *et al.*, 1999), as have the

economic costs associated with environmental damage and control (Pimentel *et al.*, 2002).

Many factors affect the establishment and spread of non-native species. These include the interaction of multiple environmental variables, such as elevation, precipitation, and soil type, which constitute the species' fundamental niche (Hutchinson, 1957; Pysek *et al.*, 2003). Non-native species have also been associated with areas of disturbance, either natural (e.g. fire or flooding; Rejmánek, 1989; Mack & D'Antonio, 1998) or human related (Macdonald *et al.*, 1988; Cowie & Werner, 1993; Gerlach *et al.*, 2003), and also influenced by abiotic factors, such as historical land use and management (Mack *et al.*, 2000).

Ideally, regular systematic field surveys should be conducted to provide the necessary information on the spread of non-native plant species. However, because of time, personnel, and financial limitations, it is not always practical or realistic to collect these primary data, and management programs may need to be based on data from previous studies. In addition, given the rapid rates of invasion, land managers would benefit greatly from the development of efficient and innovative tools, such as predictive models. A variety of models are available for predicting plant invasions. Higgins & Richardson (1996) categorized three types of invasion models based on their input requirements, outputs, and data sources: simple-demographic, spatial-phenomenological, and spatial-mechanistic models (e.g. Higgins *et al.*, 2000; Grotkopp *et al.*, 2002). Furthermore, spatially explicit models can be generated to predict the potential distribution of those species that pose the greatest threat to the ecosystem, for example, to native plant diversity (Higgins *et al.*, 1999). This would allow targeted field sampling for developing and implementing inventory and monitoring programs, and prioritizing control efforts (e.g. with respect to local 'hotspots' of native diversity; Stohlgren *et al.*, 1999).

Predictive models seek to establish a statistical relationship between environmental characteristics associated with a species occurrence to identify other, currently unsampled, areas that possess similarly suitable habitats (see Franklin, 1995; Guisan & Zimmerman, 2000; Kriticos & Randall, 2001). Franklin (1995) divided predictive modelling methods into three broad categories: Boolean, parametric models, and machine-learning methods or nonparametric methods, which vary in the assumptions they make about the underlying data distribution and whether they are discrete or probabilistic. For example, Generalized Linear Models (parametric) and Generalized Additive Models (non-parametric) (Yee & Mitchell, 1991; Brown, 1994; Austin & Meyers, 1996) demonstrate flexibility in handling nonlinear data which makes them well suited to ecological analysis that is poorly represented by classical Gaussian distributions (Guisan *et al.*, 2002). Alternatively, Classification and Regression Trees (nonparametric) offer a useful alternative when independent variables are suspected of interacting in a nested hierarchical fashion (Franklin, 1995).

This study tests the success of a machine-learning, genetic algorithm approach. Genetic algorithms were inspired by the concept of evolution by natural selection, where solutions to problems are evolved in a stochastic, iterative fashion (Holland, 1975). We apply a model called the Genetic Algorithm for Rule-set Production (GARP). GARP searches iteratively for non-random correlations between species presence and absence and environmental parameter values using four different types of rules: atomic, BIOCLIM, range rules, and logistic regression (Stockwell, 1999). Each rule is a set of if-then statements used for making inferences about the values of the variable of interest (see <http://biodi.sdsc.edu/Doc/GARP/Manual/manual.html> for a full description). The rules vary in complexity; for example, atomic rules use only a single value to predict the variable, while BIOCLIM rules contain a range of values to predict a broader environmental envelope for the species. The approach incorporates many generations of rule modification, the accuracy of

which are evaluated based on a sampling of the species data compared to a similar number of points selected randomly from the study region, and either rejected or incorporated (Stockwell & Noble, 1992).

GARP has been applied to a variety of species and geographical locations (e.g. Godown & Peterson, 2000; Anderson *et al.*, 2002; Lim *et al.*, 2002). Several advantages that have been reported for the GARP approach include:

- Genetic algorithms are a nonparametric method, thus make no assumptions about the underlying data distribution, which is often advantageous when using ecological data.
- GARP uses a rule-based approach to modelling. Each rule has preconditions that determine when it can be applied. When these conditions are not met, the rule is not used (Stockwell, 1999).
- Predictive accuracy is maximized by choosing from a variety of rule sets while at the same time minimizing rule complexity, thereby avoiding over-fitting the distribution (Kriticos & Randall, 2001). This is important when occurrence data has been collected with a less than optimal design (Peterson & Cohoon, 1999).
- Because GARP uses a combination of approaches, e.g. logistic regression, bioclimatic envelope rules, it should have a greater predictive ability than any one method applied independently (Stockwell & Peters, 1999).

The predictive modelling that we undertook for Yosemite National Park differed in two ways from previous applications of GARP. First, prior studies have generally applied GARP at continental or large regional scales using coarse resolution environmental data (4.5 km or 50 km pixels). Second, previous applications have utilized historical presence or absence data from museum specimen records. This study was conducted at a regional scale (using 30 m resolution data layers in the Geographic Information System) and used species abundance and cover data collected in 1998/99.

The objective of this study was to develop a landscape-scale predictive model of the occurrence of groups of non-native species in Yosemite National Park, California (YNP). Our strategy was to identify key environmental factors that were correlated with the percentage cover of non-native species, and to use GARP to identify other areas within the park with similar characteristics, which are presumably also vulnerable to invasion. Although these predictions will not prevent non-native plant invasions, they provide valuable information for land managers to design targeted field surveys for monitoring and determining the allocation of scarce resources.

STUDY SITE

Yosemite National Park (YNP) was designated a National Park in 1890, and has become known throughout the world because of its geological and ecological features (Fig. 1). The climate at YNP is characteristically Mediterranean; summers are warm and dry with temperatures in the Yosemite Valley (1219 m) reaching highs of 32 °C with lows of 10 °C. Winter temperatures in the Valley range from -1 °C to 10 °C. Most precipitation occurs in January, February, and March, with mean annual precipitation

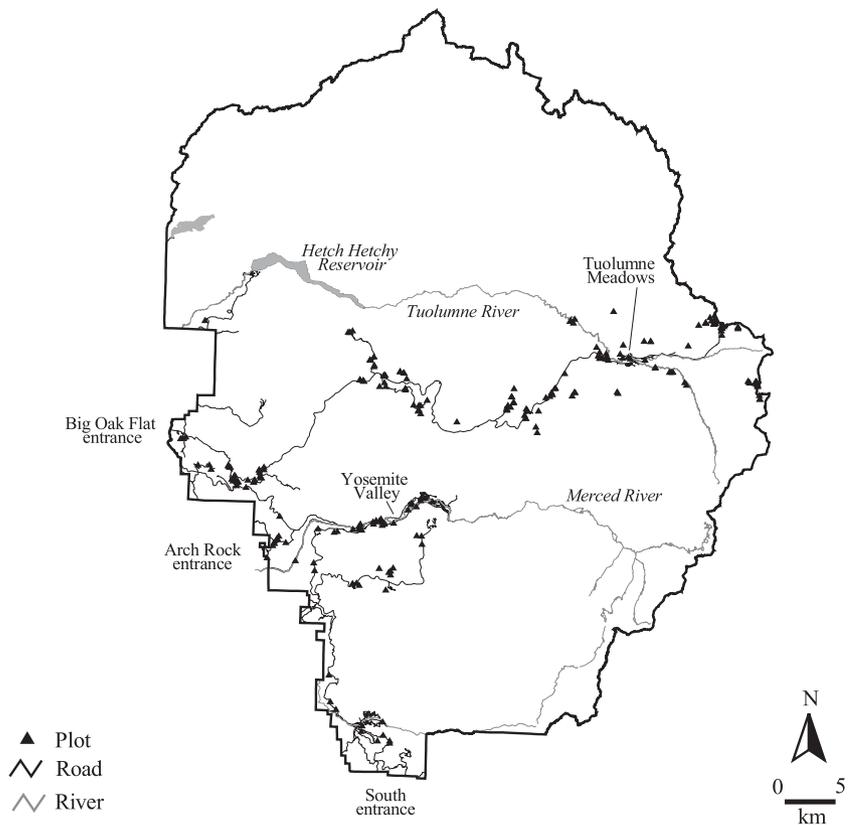


Figure 1 Yosemite National Park, California, USA, showing location of plot data used for training the Genetic Algorithm for Rule-set Production model.

ranging from 900 mm at lower elevations to 1200 mm at upper elevations. Elevation within the 3000 km² park ranges from 524 m on the western side to 3998 m in the east (Botti, 2001). The park is characterized by plutonic igneous rocks — most notably granite, and past glacial activity dominates the landscape. The park received over 3.3 million visitors in 2002, mostly concentrated in the Yosemite Valley.

Because of its size and elevation range, the park contains a wide variety of plant communities, including alpine meadows, upper montane forests, giant sequoia groves, foothill woodlands, and chaparral. These habitats support over 1400 plant taxa, including a number of endemic, threatened, or sensitive species (Botti, 2001). YNP also harbours most of the natural processes characteristic of ecosystems within the Sierra Nevada, albeit with different properties than prior to the arrival of Euro-Americans 150 years ago (e.g. fire regimes).

METHODS

Non-native species data and analyses

The data used for this study were generated from a vegetation community classification and mapping program conducted by The Nature Conservancy, the National Park Service & the US Geological Survey in 1998 and 1999 (NatureServe, 2003). The fieldwork consisted of 236 plots scaled to the physiognomy of the vegetation (herbaceous, shrub, and forested plots were 0.01 ha, 0.04 ha, and 0.1 ha, respectively) and located in all major

vegetation formations throughout the park. Plots were stratified across seven environmental variables that affect the distribution of plant species: elevation (six classes), slope (< 3°, 3–35°), aspect (Flat, NE, SE, SW, NW), geology (four classes), and fire history (burnt, unburnt in over 70 years). A cost surface was created to maximize fieldwork efficiency; permanent watercourses and very steep slopes were avoided, while distance from roads and trails was included in cost considerations. While we acknowledge that the data is not optimal for modelling purposes, the distribution of the sampling points across a variety of habitat types and the range of vegetation variables collected made it a useful dataset for designing this regional model.

At each plot, data were collected on species identity and a visual estimate of percent cover for each species. Vegetation structure variables such as the overall cover of trees, shrubs, and herbaceous species were also estimated visually for each plot (using six canopy cover classes (percent)). Plot locations were imported into a Geographic Information System (GIS) and coded with values from additional spatial data layers, providing a total of 11 environmental variables for analysis (Table 1).

Logistic regression (Hosmer & Lemeshow, 1989) was used to select the subset of the 11 environmental variables (independent variables) that best predicted the occurrence (presence/absence) of non-native species. Multiple linear regression was then used to analyse the correlation between this subset of independent variables and the species richness and cover of non-native species.

Rather than develop a separate spatial model for each of the 41 non-native species, we grouped species that were likely to: (1)

Table 1 Environment-related variables for 236 plots sampled in 1998 and 1999 in Yosemite National Park (YNP), California. Data provided by Yosemite National Park GIS group

Variable	Source	Format	Resolution
Slope (degree)	US Geologic Survey DEM	Grid	30 m
Aspect* (degree)	Derived from DEM	Grid	30 m
Elevation (meters)	Derived from DEM	Grid	30 m
Tree Cover (%)	Field	NA	NA
Shrub Cover (%)	Field	NA	NA
Herbaceous Cover (%)	Field	NA	NA
Number of Tree Species (Number ha ⁻¹)	Species Data (Field)	NA	NA
Number of Shrub Species (Number ha ⁻¹)	Species Data (Field)	NA	NA
Number of Herbaceous Species (Number ha ⁻¹)	Species Data (Field)	NA	NA
Vegetation Formation (class)	1937 YNP vegetation map	Vector	15–25 acres (6–10 ha)
Vegetation Alliance (class)	1937 YNP vegetation map	Vector	15–25 acres (6–10 ha)

*Aspect was calculated as degrees deviance (± 180) from true north for analysis.

tolerate similar environmental conditions (2) would allow land managers to capture more than one species with monitoring and control efforts, and (3) reduce computational time by limiting the number of GARP models to be run. Classification of non-native species assemblages was conducted with Two-Way Indicator Species Analysis (TWINSPAN; Gauch, 1982), using the software PC-ORD (McCune & Mefford, 1999). TWINSPAN was considered the most appropriate classification technique since we were interested in non-native species assemblages as well as the site characteristics where particular assemblages occurred. Five cut levels of abundance for separating species were specified: 0, 2, 5, 10, and 20% (since the maximum cover value was 37%), and the minimum number of species in each group specified as six. Multiple linear regression was used to analyse the relationship between the ordering of the plots and three topographic variables; elevation, slope angle, and slope aspect. A residual analysis indicated that all assumptions of linear regression were met, and no data transformations were needed.

Canonical Correspondence Analysis (CCA; ter Braak, 1995) was used to ordinate non-native species and plots along gradients of topography (elevation, slope, aspect) and vegetation structure (percent cover of trees, shrubs, and herbaceous species). To reduce the undue influence of uncommon species, those that occurred in only one plot were made 'passive' (species weight = 0; ter Braak & Smilauer, 2002). Forward stepwise multiple regression was used to select the variables that contributed significantly to the ordination, and Monte Carlo permutation tests ($n = 499$ permutations) were used to test the significance of the regression tests, the first ordination axis, and the overall ordination (ter Braak, 1995). Cover values were transformed in the species matrix ($\log Y + 1$, where Y = the original cover value). The CCA was conducted with the software Canoco 4.5 (ter Braak & Smilauer, 2002).

Predictive modelling steps

Plots were assigned values from GIS data layers which were bounded by the spatial extent of the park (Table 1). Most data

had a resolution of 30 m which, although finer resolution data would have been preferable, have the advantage that they are both cheaply and easily available, and have a long history of use in regional scale modelling (e.g. Stoms *et al.*, 1992).

Each group of species from the TWINSPAN analysis was sorted by the percent of non-native cover per plot. The 80% of plots with highest non-native cover were coded as training plots for running GARP, leaving 20% for verifying prediction results. The geographical locations of plots containing highest cover of non-native species for a particular group were specified as input data for GARP. These were then correlated against the environmental data layers shown in the CCA and linear regression analyses to be most significantly correlated with the distribution and abundance patterns of non-native species. For each of the species groups modelled in GARP, rule sets were generated for 1000 iterations each, using 50% of the input data for training and the remainder for validating each rule set, for a total of 20 times. The mean of each of the 20 runs was calculated and then imported into ArcView GIS (version 3.2) to display the probability of occurrence of each species group across the park. The final prediction surface of potential invasions was derived by adding the results of the four models from each group.

RESULTS

Analysis of non-native species distribution and abundance

A total of 41 non-native species were recorded in the plots, all of which were herbaceous (Table 2). Non-native species occurred in 23.7% of the plots ($n = 57$). The mean number of non-native species per plot was 0.69 (95% confidence intervals 0.47–0.91) with mean total cover of 2.8% (95% confidence intervals 1.4–4.2).

The occurrence (presence-absence) of non-native species had significant odds ratios for three variables; elevation, slope, and the number of herbaceous species. These variables accounted for a significant proportion of the variation in the occurrence of non-native species (McFadden's $\rho^2 = 0.515$, $P < 0.0001$). Plots

Table 2 Groups of non-native species based on TWINSpan results of sites collected in Yosemite National Park, California, during 1998 and 1999. No Defined Group indicates species with low cover values and/or restricted distribution that did not have clear group membership (see text for explanation)

Group	Species	Acronym
Group 1	<i>Aira caryophyllea</i> L.	AICA
	<i>Bromus arenarius</i> Labill.	BRAR3
	<i>Bromus diandrus</i> Roth.	BRDI10
	<i>Bromus hordeaceus</i> L.	BRHO
	<i>Bromus japonicus</i> Thunb. ex Murr.	BRJA
	<i>Bromus rubens</i> L.	BRRU2
	<i>Bromus tectorum</i> L.	BRTE
	<i>Centaurea cyanus</i> L.	CECY2
	<i>Cerastium glomeratum</i> Thuill.	CEGL2
	<i>Galium parisiense</i> L.	GAPA5
	<i>Hypericum perforatum</i> L.	HYPE
	<i>Phalaris aquatica</i> L.	PHAQ
	<i>Poa compressa</i> L.	POCO
	<i>Torilis arvensis</i> (Huds.) Link	TOAR
	<i>Trifolium repens</i> L.	TRRE3
	<i>Vulpia myuros</i> (L.) K.C. Gmel.	VUMY
Group 2	<i>Cirsium vulgare</i> (Savi) Ten.	CIVU
	<i>Citrullus lanatus</i> var. <i>lanatus</i> (Thunb.) M. & N.	CILAL
	<i>Myosotis discolor</i> Pers.	MYDI
	<i>Poa pratensis</i> L.	POPR
Group 3	<i>Taraxacum officinale</i> G.H. Weber ex Wiggers	TAOF
	<i>Dianthus armeria</i> L.	DIAR
	<i>Holcus lanatus</i> L.	HOLA
	<i>Lactuca serriola</i> L.	LASE
Group 4	<i>Leucanthemum vulgare</i> Lam.	LEVU
	<i>Tragopogon dubius</i> Scop.	TRDU
	<i>Agrostis gigantea</i> Roth	AGGI2
	<i>Erodium cicutarium</i> (L.) L'Her. ex Ait.	ERCI6
No Defined Group	<i>Phleum pratense</i> L.	PHPR3
	<i>Rumex acetosella</i> L.	RUAC
	<i>Rumex crispus</i> L.	RUCR
	<i>Agrostis capillaris</i> L.	AGCA5
	<i>Agrostis stolonifera</i> L.	AGST2
	<i>Bromus sterilis</i> L.	BRST2
	<i>Plantago lanceolata</i> L.	PLLA
	<i>Poa annua</i> L.	POAN
	<i>Poa nemoralis</i> L.	PONE
<i>Rubus discolor</i> Weihe & Nees	RUDI2	
<i>Silene gallica</i> L.	SIGA	
<i>Vinca major</i> L.	VIMA	

with non-native species tended to occur most frequently at relatively low to mid-elevations in flat areas with other herbaceous species. The classification success of plots with non-native species was 65%, and the classification success of plots without non-native species was 89%. The overall classification success was 83.5%. Multiple linear regression analysis indicated that non-native species richness and cover both had significant negative correlations with elevation and the number of herbaceous species, but no significant relationship with slope.

TWINSpan of the plot data indicated four major groupings of non-native species (Table 2), with one or two species in each accounting for most of the cover. The largest group was dominated by *Bromus tectorum* and *Vulpia myuros* (group one). Group two was dominated by *Poa pratensis* and *Cirsium vulgare*.

The third group was dominated by *Holcus lanatus*, and the fourth by *Rumex acetosella*. A fifth group consisting of uncommon species such as *Poa annua*, *Silene gallica*, and *Rubus discolor* had no clear relationships among themselves or with the other species.

The multiple regression of the topographic variables with the ordering of the plots was significant ($R = 0.622$, $F = 10.30$, d.f. = 3.49, $P < 0.001$). Elevation had a significant positive relationship and slope angle had a significant negative relationship with the ordering of the plots (Table 3). In general, plots characterized by *Bromus tectorum* and *Vulpia myuros* tended to occur in steep sites at lower elevations, while plots characterized by *Poa pratensis* and *Cirsium vulgare* were flatter and at higher elevations. The other two species groups tended to be in plots that were intermediate in elevation and moderate in steepness.

Table 3 Multiple linear regression statistics of the relationship between three topographic variables and the TWINSPLAN ordering of 57 plots in Yosemite National Park, California, 1998–99

Variable	Coefficient	SEM Coefficient	<i>T</i>	<i>P</i>
Constant	11.55	7.09	1.63	0.110
Slope	-1.035	0.286	3.62	0.001
Elevation	0.015	0.004	4.02	0.001
Aspect	0.022	0.032	0.69	0.495

Adjusted $R^2 = 0.386$, d.f. = 3,49, $P = 0.001$.

Table 4 Results of forward stepwise multiple regression of 6 variables for a Canonical Correspondence Analysis of 41 non-native species in 57 plots in Yosemite National Park, California, 1998–99

Variable	Variance Explained	<i>F</i>	<i>P</i>
Shrub Cover (%)	0.47	4.30	0.002
Herbaceous Cover (%)	0.23	2.24	0.010
Tree Cover (%)	0.23	2.13	0.018
Elevation (m)	0.16	1.46	0.102
Slope Angle (°)	0.14	1.41	0.152
Aspect	0.11	1.04	0.394

Table 5 Results of Canonical Correspondence Analysis of 41 non-native species in 57 plots in Yosemite National Park, California, 1998–99. λ = eigenvalue for the axis

CCA Statistic	Axis 1	Axis 2
λ	0.534	0.227
Species-environment correlations	0.875	0.763
Cumulative percentage variance of species-environment relation	57.1	81.4
Variable		
Tree Cover (%)	-0.374	0.179
Shrub Cover (%)	-0.797	-0.079
Herbaceous Cover (%)	-0.458	-0.529

Fifteen species occurred in only one plot, therefore they were held passive in the CCA. Forward stepwise multiple regression indicated that three of the environmental variables had significant correlations with the distribution of non-native species: the percent cover of trees, shrubs, and herbaceous species (Table 4). Monte Carlo permutation tests indicated that there was a significant difference in the distribution of non-native species for the first axis ($\lambda_1 = 0.534$), the second axis ($\lambda_2 = 0.227$), and the overall ordination in the CCA ($P = 0.002$; Table 5). The first two axes explained 81.4% of the cumulative variation in non-native species distribution (Table 5). The relatively strong species-environment correlation (the linear combination of environmental variables and sample scores based on the species data) on the first axis indicated that this was primarily a gradient from plots on steep slopes dominated by shrubs to plots on flatter

terrain dominated either by trees or in open meadows (Table 5 and Fig. 2). The species-environment correlation for the second axis was moderately strong, and interpreted as a gradient from forests to meadows (Table 5 and Fig. 2).

Based on these results we determined that five variables would be included in the GARP models; three vegetation structure variables from the CCA (percent cover of trees, shrubs, and herbaceous species) and two topographic variables from the regression analyses (slope and elevation). Because the GIS data layers did not have specific estimates of cover for the vegetation structure layers, we used vegetation alliance as a substitute.

Predictive modelling

A GARP model was run for each of the four species groups. Model results ranged from 0 to 1; a value of one implied that a particular location (grid cell) had the highest probability (100%) of an occurrence for that species group, i.e. it contained a combination of environmental data layers most similar to that of the input plots. A value of zero meant that it had no probability of a non-native species occurrence.

The predicted distribution of the species within group one was largely driven by elevation (Fig. 3a). The species, which include *Bromus tectorum* and *Vulpia myuros*, were predicted to occur in areas less than 1600 m, which would be expected for these largely California annual grassland species. The preference of these species for flat areas is illustrated by their distribution eastwards along the riparian areas of the Tuolumne and Merced Rivers and in the Yosemite Valley. These results reflect the published literature; *V. myuros* is limited to elevations less than 1300 m in California, while *B. tectorum* is known to prefer non-forested areas (e.g. open burnt areas or meadows) (Tirmenstein, 1987). Concentrations of the species are predicted in the Yosemite Valley and the South, Big Oak Flat, and Arch Rock park entrances, which receive high amounts of vehicle and human traffic.

The species in group two, dominated by *Poa pratensis* and *Cirsium vulgare*, had the widest predicted distribution (Fig. 3b). Species are predicted to occur in areas less than 2700 m but not in steeply sloping terrain ($< 13^\circ$), and across a range of vegetation alliances. Again, this is consistent with the published literature on these species; *C. vulgare* is known to tolerate dry to moist habitats, with little preference for soil type providing there is intermediate moisture, and does poorly on steep slopes (Randall, 1990; Hickman, 1993). *P. pratensis* also characterizes numerous vegetation types at moderate to high elevations and, again, has a wide tolerance of different soil types (Uchytel, 1993). Within YNP, *C. vulgare* has been observed extensively in burnt areas across a range of elevations (P. Moore, pers. obs.) and *Taraxacum officinale*, another species in this group, has been noted in many high-elevation meadows and roadsides in the southern Sierra Nevada (S. Haultain, pers. comm.). Predicted concentrations of these species occur in areas with a high number of trails, campgrounds, and heavy visitation, e.g. the Yosemite Valley and the South entrance of the park, while predicted occurrences include the Tuolumne Meadows area.

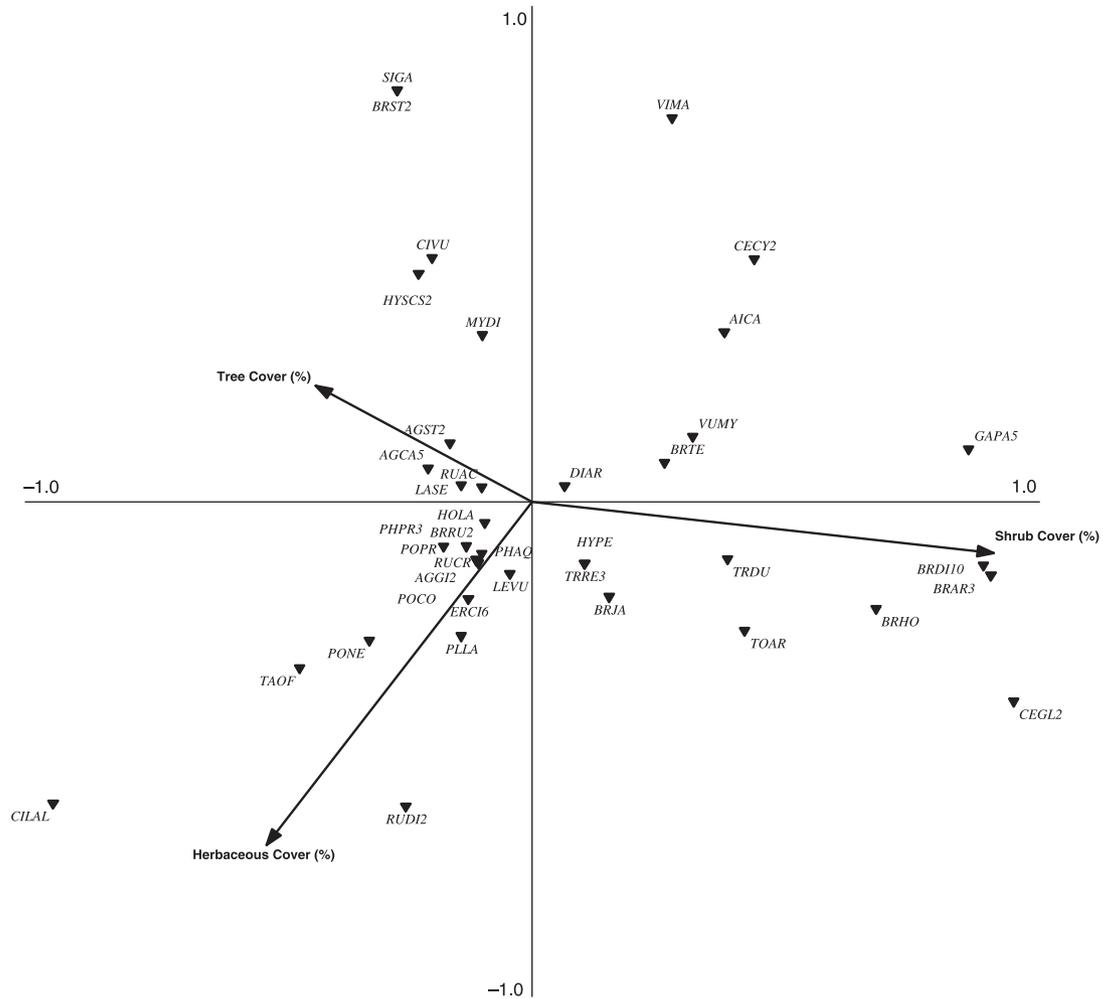


Figure 2 Canonical Correspondence Analysis triplot of the first two ordination axes for 57 0.1-ha plots where non-native species ($n = 41$) were present in Yosemite National Park, California. Species acronyms are given in Table 2.

The species in group three, notably *Holcus lanatus*, are largely restricted to narrow, flat, corridors along the lowest elevations of the Tuolumne, Merced, and South Fork of the Merced rivers (Fig. 3c). This distribution pattern strongly follows the ponderosa pine and foothill pine-live oak-chaparral woodland alliances in the vegetation data. Finally, the species in group four, dominated by *Rumex acetosella*, are predicted across a broad range of environmental conditions (Fig. 3d). *R. acetosella*, for instance, is reported to tolerate elevations up to 2200 m, prefer open meadows, and is adaptable to a variety of different soil types (Wilson & Tilman, 1991). Predicted occurrences form a wide swath along the western edge of the park, with highest predictions at the South entrance of the park and in the Yosemite Valley.

The results of the four models were added to give a final prediction landscape for non-native species occurrences (Fig. 4). At a coarse scale, the accuracy of the model can be assessed by looking at the number of field plots (the 20% omitted from the model) that were accurately predicted as ‘present’ compared to the number that were not captured (‘absent’). Correct prediction of non-native species presence for the model was 76%. The

prediction results were further analysed by assessing the probability with which these points were predicted. If the model was completely accurate, all of the validation points would have a probability of 100%, since we know from the field data that these plots harbour non-native species. The model successfully predicted a proportion of the plots with 100% accuracy, and also a notable number with higher than 75% probability (Fig. 5). However, the model failed to predict (i.e. missed) 24% of the validation plots.

DISCUSSION

For each of the four species groups a GARP model was run using the most significant environmental data layers as determined by the community analyses: elevation, slope, and vegetation structure variables. The 75% validation result from the final model is indicative of the potential that the GARP model has for predicting sites where non-native species could invade. These results, however, must be interpreted with caution since the data were not collected specifically for modelling; e.g. the number and

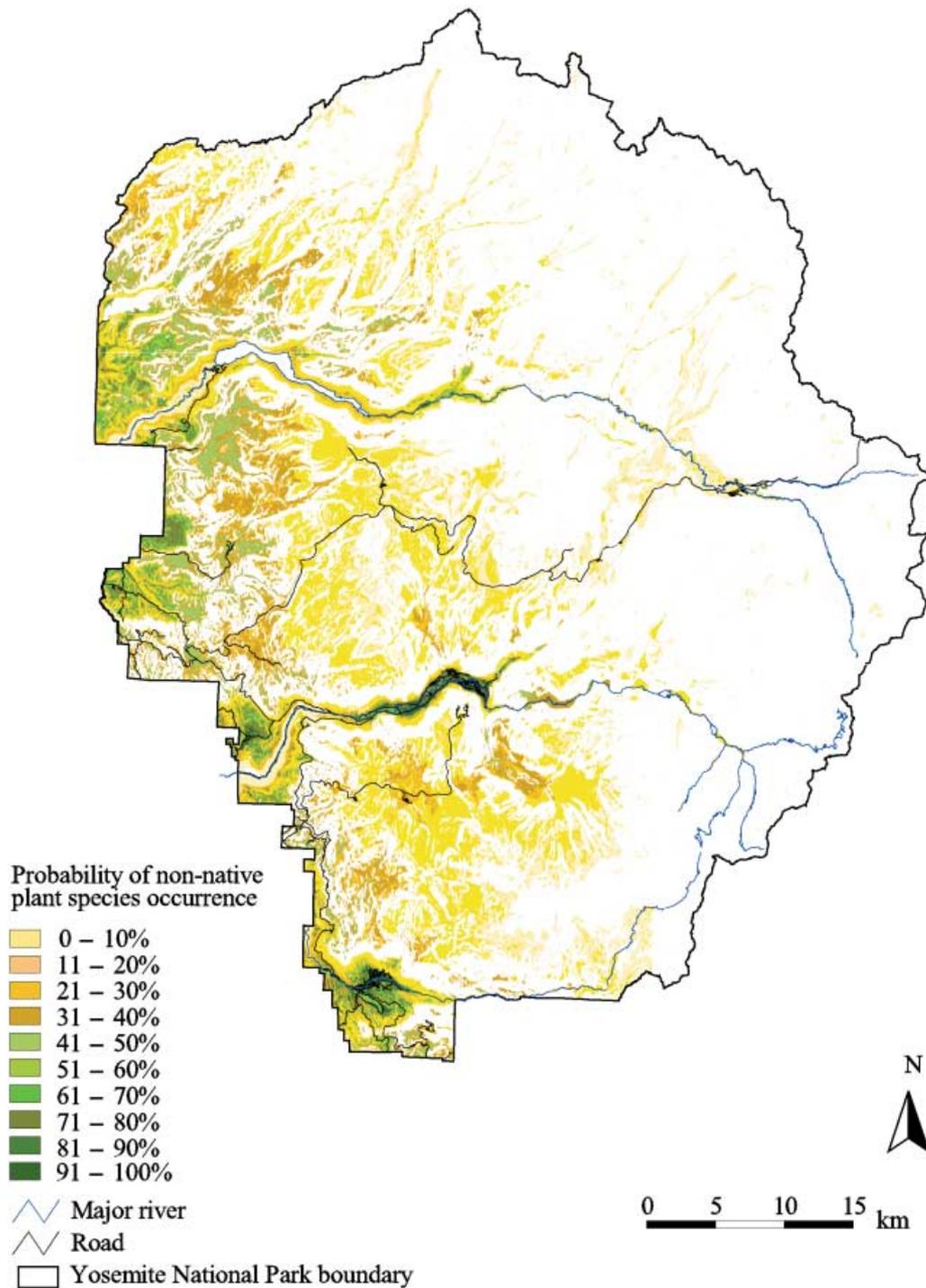


Figure 4 Predicted occurrences of all four species groups based on site data collected in 1998–99 and environmental data layers, Yosemite National Park, California.

variables which provided the justification for using these as inputs in the predictive model. Even though cover values were relatively low, patterns between species and variables were qualitatively consistent between the different analyses and represent legitimate ecological patterns and relationships. Elevation and slope angle can be considered indirect gradients in that these factors do not have a direct physiological effect on

the species (Austin, 2002), but do exert an influence through their correlation with temperature and soil moisture (direct gradients). The preference of non-native species for lower elevations largely reflects their range in their native habitats (Hickman, 1993), which assists in explaining the predicted swath of non-native species in the lower elevation, western portion of the park.

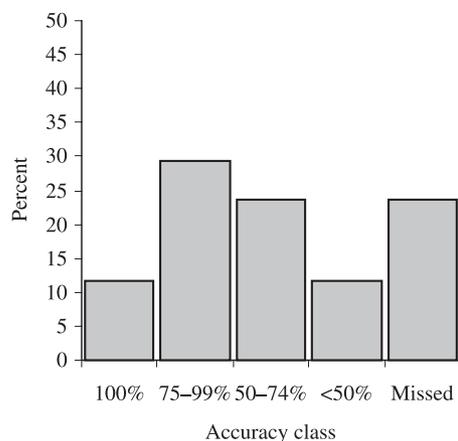


Figure 5 Accuracy assessment of final predictive model for all species groups: percent of validation plots predicted at different levels of accuracy using data from sites collected in 1998 and 1999, Yosemite National Park, California.

The plot ordering in the TWINSPLAN implied that non-native species composition changed from steep, low-elevation hillsides where soil moisture might be low, to flatter, mid-elevation sites (such as meadows) where soil moisture was relatively high. This is consistent with the fact that these species were all herbaceous, preferring areas with higher moisture and light. Although non-native species can invade less mesic sites (Brooks & Pyke, 2001), there appears to be lower rates of invasion into communities in more extreme environments (Rejmánek, 1989). A further consideration is that results might have differed substantially if the 41 species had been modelled separately, but our intention was to group similar species that might exist in similar environmental conditions, allowing management to efficiently monitor and control like species.

The predicted occurrences of non-native species in YNP can also be viewed in relation to disturbances, both natural and human-related. The western portion of the park, particularly the southern and western entrances, and the Yosemite Valley were predicted to have moderate to high levels of non-native plant occurrences. This raises concern because these areas have high levels of human activity and disturbance. For example, the Valley receives the majority of the park's visitors and, as does the South entrance, contains numerous campgrounds, roads, and trails. The predicted occurrence of species within group two in the Tuolumne Meadows area is particularly concerning, given the number of trailheads which access remote areas in wilderness.

The levels of human use in these areas provide opportunities for spreading non-native species via human and vehicle vectors and for potentially dispersing seeds by foot into remote areas. The concentration of some species along stream corridors is also important because propagules can be transported by annual spring floods and periodic winter floods to riparian habitats which are important to wildlife and are among the most species-rich vegetation types in the Sierra Nevada. The location of non-native species in predominantly natural and human disturbed habitats is particularly critical given the ability of some non-

natives to exploit a pulse of resources, e.g. *Cirsium vulgare* (Forcella & Wood, 1986). Furthermore, these vulnerable areas at lower elevations might also represent staging areas where species could slowly adapt to higher elevation conditions.

Shortcomings of predictive modelling

While GARP has certain advantages to other modelling approaches, some potential shortcomings include:

- The stochastic nature of genetic algorithms (e.g. mutation, Stockwell & Noble, 1992) mean that multiple solutions can be provided with the same optimization criterion (Anderson *et al.*, 2002), which necessitates careful evaluation or further processing.
- The predictive accuracy (rule fitness) is tested by comparing points resampled from known occurrences and from the background (Stockwell & Peters, 1999). Consequently, some data are excluded and the algorithm cannot take advantage of all locality records (Anderson *et al.*, 2002).
- GARP uses only presence data because absence data is not always available, which causes asymmetry (i.e. errors in pseudo-absences) in the input data (Anderson *et al.*, 2002).

Genetic algorithms also suffer from shortcomings of other predictive modelling techniques. First, predicted distributions may represent overpredictions by identifying potential habitat to which species are unable to disperse (Peterson & Vieglais, 2001). Alternatively, models may underpredict the fundamental niche (as in this study) owing to limited data layers, or explanatory variables (Anderson *et al.*, 2002), and minimal occurrence data for training the model.

Second, predictive models are often static and fail to reflect the stochastic nature of the landscape, assuming species are in equilibrium with their environment (Guisan & Theurillat, 2000; Guisan & Zimmerman, 2000). The incorporation of ecological processes, such as fire history or flood regimes in riparian areas, might have greater explanatory power in determining species distributions (Austin, 2002). In addition, particularly for this study site, the integration of anthropogenic-related disturbances would be valuable, given the positive correlation between the number of invasive species and road and population density in California (Dark, 2004). This is particularly important for invasive species which have been associated with areas of disturbance. Furthermore, predictive models often fail to incorporate information on propagule pressure which determines the ease with which invasive plants overcome environmental barriers to become established (Rouget & Richardson, 2003). The incorporation of propagule pressure, along with environmental variables, could assist in predicting the percentage cover of invasive species (Rouget & Richardson, 2003).

Third, predictive models often assume a symmetrical, bell-shaped response of a species to environmental variables (Austin, 1999, 2002). However, evidence for this response shape is weak. For example, Austin *et al.* (1994) found a significantly skewed realized niche response of eucalypt trees to temperature in south-eastern temperate forests of Australia.

Fourth, predicted distributions of species are often based on a limited number of plots within a discretely defined area.

Findings cannot be extrapolated beyond this area since correlations between species occurrences and topographic variables (elevation, slope, aspect) result in different combinations of direct and resource gradients in other regions (Guisan & Zimmerman, 2000). For modelling at broader spatial scales, additional occurrence data across the range of explanatory variables would be required, and perhaps complimented with information gleaned from published literature on the species. Furthermore, in this study, there is an obvious disjunction between the resolution of the input data layers compared to the scale at which non-native species invasions occur.

Fifth, from a temporal-scale perspective, YNP is in the early stages of invasion. In the context of California, the bioregion in which YNP is located harbours 16–30 invasive non-native plants, compared to 45–59 in more heavily invaded coastal habitats (Dark, 2004). Classification success might be higher had non-native species reached the full extent of habitat that could support them. This is a particularly important consideration with invasive plants, which might exhibit increasing rates of dispersal following disturbance. For example, Keeley *et al.* (2003) found time since fire was generally the most critical factor in non-native invasions in chaparral sites in the Sierra Nevada. Non-natives were increasingly recorded in later successional years once propagules had time to colonize the postfire, open canopy shrublands (Keeley *et al.*, 2003).

Future predictive modelling directions and implications for management

Three tasks would assist in improving the predictive model for managing non-native species in YNP. First, conducting a rapid ground validation of results to confirm occurrences and provide cover estimates of non-natives at the predicted locations. Second, mapping locations of non-native plants would allow further refinement of the model by weighting areas based on proximity to sources of propagules. Third, we recommend performing experimental studies to explore the cause of the statistical correlations between species and the environmental variables (Mack, 1996; Austin, 2002). For example, investigating invasion patterns along disturbance gradients, such as distance to road, or conducting site-specific experiments to identify direct gradients and proximal variables (Austin, 2002). Fourth, results from control efforts can feed into an adaptive management framework to evaluate the success of non-native species control efforts over the long-term.

The need to develop predictive models to determine the potential distribution of non-native species has been recognized (Hobbs & Mooney, 1998). This need is vital given the ongoing and expected increase in the ecological and economic impacts of non-native species in national parks and other nature preserves. The objective of this study was to produce spatial results to assist management in decision-making to encourage a more proactive approach for managing non-native species.

First, the maps of predicted occurrences will increase the effectiveness of inventory and monitoring activities of non-natives at YNP; results can be used to efficiently target monitoring to areas

more likely to contain non-natives. For example, monitoring remote areas where non-natives might establish a foothold and otherwise go unnoticed, such as remote areas burnt by wild fire. Second, control efforts can be prioritized according to the likelihood of non-native species occurrence; the Yosemite Valley, South entrance of the park, and Tuolumne Meadows (an area surrounded by low-probability landscape that would otherwise not be considered vulnerable). The identification of species that tolerate similar environmental conditions implies they may respond to similar control efforts, which could lead to a more efficient application of limited resources. Third, these results can feed into other management decisions, such as postfire non-native plant management, which can be modified to the vulnerability level of the area burnt.

CONCLUSIONS

The Vail Agenda's strategic plan for the National Park Service in 1991 emphasized the importance of good quality information on natural resources in the National Park system being readily available to park managers (Babbitt, 1992). The research detailed in this paper assists Yosemite National Park in addressing this goal by helping to stratify inventory, monitoring, and control activities in relation to non-native species. The broader implications of this research are that similar predictive models can be developed for other national parks where suitable plot data exist, or to other conservation issues such as rare plant distributions, to increase the efficiency and reduce the costs of field work. With the expected increase in the spread of non-native plants, predictive modelling techniques offer a valuable tool for planning control and monitoring activities of non-native plant populations.

ACKNOWLEDGEMENTS

The authors thank Linda Mutch from Sequoia and Kings Canyon National Parks for assistance. Marcel Rejmánek, Dave Vieglais, Eva Grotkopp, and Jennifer Erskine for helpful review comments, and Jim Quinn for logistical support. Funding, spatial data, and plot data were supplied by the National Park Service.

REFERENCES

- Anderson, R.P., Gomez-Laverde, M. & Peterson, A.T. (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography*, **11**, 131–141.
- Austin, M.P. (1999) A silent clash of paradigms: some inconsistencies in community ecology. *Oikos*, **86**, 170–178.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Austin, M.P. & Meyers, J.A. (1996) Current approaches to modelling the environmental niche of eucalypts: implications for management of forest biodiversity. *Forest Ecology and Management*, **86**, 95–106.

- Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. (1994) Determining species response functions to an environmental gradient by means of a β -function. *Journal of Vegetation Science*, **5**, 215–228.
- Babbitt, B. (1992) *National parks for the twenty-first century: the Vail Agenda*. Report and recommendations to the director of the National Park Service. National Park Foundation, Washington, D.C., USA.
- Botti, S.J. (2001) *An illustrated flora of Yosemite National Park*. The Yosemite Association, El Portal, California, USA.
- Brooks, M.L. & Pyke, D.A. (2001) Invasive plants and fire in the deserts of North America. *Proceedings of the invasive species workshop: the role of fire in the control and spread of invasive species fire conference 2000: the first national congress on fire ecology, prevention, and management* (ed. by K.E.M. Galley & T.P. Wilson), pp. 1–14. Tall Timbers Research Station, Tallahassee, Florida, USA.
- Brown, D.G. (1994) Predicting vegetation types at treeline using topography and biophysical disturbance variables. *Journal of Vegetation Science*, **5**, 642–656.
- Cowie, I.D. & Werner, P.A. (1993) Non-native plant species invasive in Kakadu National Park, tropical northern Australia. *Biological Conservation*, **63**, 127–135.
- Dark, S.J. (2004) The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. *Diversity and Distributions*, **10**, 1–9.
- Drake, J.A., Mooney, H.A., Di Castri, F., Groves, R.H., Kruger, F.J., Rejmánek, M. & Williamson, M., eds. (1989) *Biological invasions: a global perspective*. John Wiley & Sons, Chichester.
- Forcella, F. & Wood, H. (1986) Demography and control of *Cirsium vulgare* (Savi) Ten. in relation to grazing. *Weed Research*, **26**, 199–206.
- Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474–499.
- Gauch, H.G. (1982) *Multivariate analysis in community ecology*. Cambridge University Press, New York, USA.
- Gerlach, J.D. Jr, Moore, P.E., Johnson, B., Roy, D.G., Witmarsh, P., Lubin, D.M., Graber, D.M., Haultain, S., Pfaff, A. & Keeley, J.E. (2003) *Alien plant species threat assessment and management prioritization for Sequoia-Kings Canyon and Yosemite National Parks*. United States Geological Survey; Open-File Report 02–170; Carson City, Nevada.
- Godown, M.E. & Peterson, A.T. (2000) Preliminary distributional analysis of endangered US bird species. *Biodiversity and Conservation*, **9**, 1313–1322.
- Gordon, D.R. (1998) Effects of invasive, non-indigenous plant species on ecosystem processes. Lessons from Florida. *Ecological Applications*, **8**, 975–989.
- Grotkopp, E., Rejmánek, M. & Rost, T.L. (2002) Toward a causal explanation of plant invasiveness: seedling growth and life-history strategies of 29 pine (*Pinus*) species. *American Naturalist*, **159**, 396–419.
- Guisan, A., Edwards, T.C. Jr & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89–100.
- Guisan, A. & Theurillat, J.-P. (2000) Equilibrium modelling of alpine plant distribution: how far can we go? *Phytocoenologia*, **30**, 353–384.
- Guisan, A. & Zimmerman, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hickman, J.C. (1993) *The Jepson manual: higher plants of California*. University of California Press, Berkeley, USA.
- Higgins, S.I. & Richardson, D.M. (1996) A review of models of alien plant spread. *Ecological Modelling*, **87**, 249–265.
- Higgins, S.I., Richardson, D.M. & Cowling, R.M. (2000) Using a dynamic landscape model for planning the management of alien plant invasions. *Ecological Applications*, **10**, 1833–1848.
- Higgins, S.I., Richardson, D.M., Cowling, R.M. & Trinder-Smith, T.H. (1999) Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conservation Biology*, **13**, 303–313.
- Hobbs, R.J. & Mooney, H.A. (1998) Broadening the extinction debate: population deletions and additions in California and western Australia. *Conservation Biology*, **12**, 271–283.
- Holland, J.H. (1975) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press, Ann Arbor, USA.
- Hosmer, D.W. & Lemeshow, S. (1989) *Applied logistic regression*. Wiley, New York, USA.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
- Keeley, J.E., Lubin, D. & Fotheringham, C.J. (2003) Fire and grazing impacts on plant diversity and alien plant invasions in the southern Sierra Nevada. *Ecological Applications*, **13**, 1355–1374.
- Kriticos, D.J. & Randall, R.P. (2001) A comparison of systems to analyze potential weed distributions. *Weed risk assessment* (ed. by R.H. Groves, F.D. Panetta & J.G. Virtue), pp. 61–79. CSIRO Publishing, Australia.
- Lim, B.K., Peterson, A.T. & Engstrom, M.D. (2002) Robustness of ecological niche modelling algorithms for mammals in Guyana. *Biodiversity and Conservation*, **11**, 1237–1246.
- Macdonald, I.A.W., Graber, D.M., DeBenedetti, S., Groves, R.H. & Fuentes, E.R. (1988) Introduced species in nature reserves in Mediterranean-type climatic regions of the world. *Biological Conservation*, **44**, 37–66.
- Mack, R.N. (1996) Predicting the identity and fate of plant invaders: emergent and emerging approaches. *Biological Conservation*, **78**, 107–121.
- Mack, M.C. & D'Antonio, C.M. (1998) Impacts of biological invasions on disturbance regimes. *Trends in Ecology and Evolution*, **13**, 195–198.
- Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M. & Bazzaz, F.A. (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689–710.

- McCune, B. & Mefford, M.J. (1999) PC-ORD. *Multivariate analysis of ecological data*. MjM Software Design, Gleneden Beach, Oregon, USA.
- NatureServe. (2003) *Classification of the vegetation of Yosemite National Park and surrounding environs in Tuolumne, Mariposa, Madera and Mono counties, California*. Report to National Park Service in cooperation with the California Native Plant Society and California Natural Heritage Program, Wildlife and Habitat Data Analysis Branch. California Department of Fish and Game. Sacramento, California, USA.
- Parker, I.M., Simberloff, D., Lonsdale, W.M., Goodell, K., Wonham, M., Kareiva, P.M., Williamson, M.H., von Holle, B., Moyle, P.B., Byers, J.E. & Goldwasser, L. (1999) Impact: toward a framework for understanding the ecological effects of invaders. *Biological Invasions*, **1**, 3–19.
- Peterson, A.T. & Cohoon, K.P. (1999) Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling*, **117**, 159–164.
- Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modelling: new approaches from bioinformatics attack a pressing problem. *Bioscience*, **51**, 363–371.
- Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2002) Environmental and economic costs associated with non-indigenous species in the United States. *Bioscience*, **50**, 53–65.
- Pyšek, P., Jarosik, V. & Kucera, T. (2003) Inclusion of native and alien species in temperate nature reserves: an historical study from Central Europe. *Conservation Biology*, **17**, 1414–1424.
- Randall, J.M. (1990) Establishment and control of bull thistle (*Cirsium vulgare*) in Yosemite Valley. *Examples of resource inventory & monitoring in national parks of California: Proceedings 3rd biennial conference on research in California's national parks* (ed. by C. Van Riper, T.J. III Stohlgren, S.D. Jr Veirs & S. Castillo Hillyer), pp. 177–193. Transactions and Proceedings Series No 8. US Department of the Interior, National Park Service, Washington, D.C., USA.
- Rejmánek, M. (1989) Invasibility of plant communities. *Biological invasions: a global perspective* (ed. by J.A. Drake, H.A. Mooney, F. Di Castri, R.H. Groves, F.J. Kruger, M. Rejmánek & M. Williamson), pp. 369–388. John Wiley & Sons, Chichester.
- Rouget, M. & Richardson, D.M. (2003) Inferring process from pattern in alien plant invasions: a semi-mechanistic model incorporating propagule pressure and environmental factors. *American Naturalist*, **164**, 713–724.
- Simberloff, D., Schmitz, D.C. & Brown, T.C., eds. (1997) *Strangers in Paradise*. Island Press, Washington, D.C., USA.
- Stockwell, D.R.B. (1999) Genetic algorithms II. *Machine learning methods for ecological applications* (ed. by A.H. Fielding), pp. 123–144. Kluwer Academic Publishers, Boston, USA.
- Stockwell, D.R.B. & Noble, I.R. (1992) Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computer Simulation*, **33**, 385–390.
- Stockwell, D.R.B. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science*, **13**, 143–158.
- Stohlgren, T.J., Binkley, D., Chong, G.W., Kalkhan, M.A., Schell, L.D., Bull, K.A., Otsuki, Y., Newman, G., Bashkin, M. & Son, Y. (1999) Exotic plant species invade hot spots of native plant diversity. *Ecological Monographs*, **69**, 25–46.
- Stoms, D.M., Davis, F.W. & Cogan, C.B. (1992) Sensitivity of wildlife habitat models to uncertainties in GIS data. *Photogrammetric Engineering and Remote Sensing*, **58**, 843–850.
- ter Braak, C.J.F. (1995) Ordination. *Data analysis in community and landscape ecology* (ed. by R.H.G. Jongman, C.J.F. ter Braak & O.F.R. van Tongeren), pp. 91–173. Cambridge University Press, New York, USA.
- ter Braak, C.J.F. & Smilauer, P. (2002) *CANOCO 4 5 users guide*. Centre for Biometry, Wageningen, Netherlands.
- Tirmenstein, D.A. (1987) *Vulpia myuros*. US Department of Agriculture. Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory. Fire Effects Information System. Available from <http://www.fs.fed.us/database/feis/> (accessed February 2003).
- Uchytel, R.J. (1993) *Poa Pratensis*. US Department of Agriculture. Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory. Fire Effects Information System. Available from <http://www.fs.fed.us/database/feis/> (accessed February 2003).
- Vitousek, P. (1986) Biological invasions and ecosystem properties: can species make a difference? *Ecology of biological invasions of North America and Hawaii* (ed. by H.A. Mooney & J.A. Drake). Springer-Verlag, New York.
- Wilson, S.D. & Tilman, D. (1991) Interactive effects of fertilization and disturbance on community structure and resource availability in an old-field plant community. *Oecologia*, **88**, 61–71.
- Yee, T.W. & Mitchell, N.D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Sciences*, **2**, 587–602.