

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/227017019>

GLM versus CCA Spatial Modeling of Plant Species Distribution

Article in *Plant Ecology* · July 1999

DOI: 10.1023/A:1009841519580

CITATIONS

444

READS

725

3 authors, including:



Antoine Guisan

University of Lausanne

321 PUBLICATIONS 36,471 CITATIONS

[SEE PROFILE](#)



Stuart Weiss

Creekside Center for Earth Observation

78 PUBLICATIONS 2,858 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



CONS_BIOREC: CONserving Biodiversity under an uncertain future by REConciling conflicting conservation perspectives: lessons from a multi-scale framework in the Iberian Peninsula [View project](#)



Microbial biogeography and global change: Linking Pyrosequencing, phylogenetics and soil fungal and bacterial communities along a wide elevation gradient in the Alps. [View project](#)

GLM versus CCA spatial modeling of plant species distribution

Antoine Guisan¹, Stuart B. Weiss² & Andrew D. Weiss²

¹*Botanical Center, University of Geneva, CP60, CH-1292 Geneva, Switzerland; and Swiss Center for Faunal Cartography (CSCF) CH-2000 Neuchâtel, Switzerland;* ²*Center for Conservation Biology, Department of Biological Sciences, Stanford University, CA 94035, USA*

Received 16 April 1998; accepted in revised form 17 December 1998

Key words: Constrained ordination, Disturbances, Logistic regression, Model comparison, Plant distribution, Spatial modeling, Spring Mountains (Nevada)

Abstract

Despite the variety of statistical methods available for static modeling of plant distribution, few studies directly compare methods on a common data set. In this paper, the predictive power of Generalized Linear Models (GLM) versus Canonical Correspondence Analysis (CCA) models of plant distribution in the Spring Mountains of Nevada, USA, are compared. Results show that GLM models give better predictions than CCA models because a species-specific subset of explanatory variables can be selected in GLM, while in CCA, all species are modeled using the same set of composite environmental variables (axes). Although both techniques can be readily ported to a Geographical Information System (GIS), CCA models are more readily implemented for many species at once. Predictions from both techniques rank the species models in the same order of quality; i.e. a species whose distribution is well modeled by GLM is also well modeled by CCA and vice-versa. In both cases, species for which model predictions have the poorest accuracy are either disturbance or fire related, or species for which too few observations were available to calibrate and evaluate the model. Each technique has its advantages and drawbacks. In general GLM will provide better species specific-models, but CCA will provide a broader overview of multiple species, diversity, and plant communities.

Abbreviations: AML – ArcInfo Macro Language; CCA – Canonical Correspondence Analysis; DEM – Digital Elevation Model; GIS – Geographical Information System; GLM – Generalized Linear Model.

Introduction

Static modeling of plant distribution using statistical methods to relate vegetation to the environment has gained importance in recent years (see Franklin 1995). A variety of statistical techniques – ordination, GLM, GAM, etc. – have been proposed and used. Here, we do not aim at giving a list of examples for these distinct approaches, which information is best provided in review papers (e.g., Franklin 1995) or books (e.g., Jongman et al. 1995). Most studies use only one of the many statistical techniques that may properly be used, and little information is available on the respective predictive capacity of each approach compared to the others. The debate is usually restricted to the

intrinsic suitability of a particular method for a given data set. For instance, is a non-normally distributed response variable still properly modeled by least-square methods or weighted averaging methods (e.g., reciprocal averaging in canonical correspondence analysis) that both rely on the normality assumption (Palmer 1993; Jongman et al. 1995)? Or, with GLMs, is a skewed response curve best integrated in a model by a cubic polynomials (e.g., Ferrer-Castán et al. 1995), a β -function (e.g., Austin et al. 1994), a non-linear function (e.g. from a hierarchical set of functions; Huisman et al. 1993) or a smoothed function (e.g., Yee & Mitchell 1991)? Again, answers to such questions can be found in specific papers (e.g., Oksanen 1997). In turn, studies that compare the results of different

techniques applied to a similar data set (as, e.g., Skidmore et al. 1996), be it real or simulated, are missing in the literature.

The aim of this paper is to compare two broadly used techniques - CCA-based modeling and GLM methods - by using the same training and evaluation data sets. In order to have a sufficient number of models to compare, models are calibrated for twenty-three species by using the same basic set of predictor variables. More specifically, our main goals are:

- (1) To compare the results (fit of the model, adequacy of the prediction) of both sets of models. Does the overall trend in the results agree for both sets? Is a well GLM modeled species also well modeled by CCA compared to models for other species in the same data set?
- (2) From the overall trend between both sets of model, is it possible to distinguish some grouping of species having a better modeling result than others? Are there ecological reasons for this (e.g., missing predictors in the data set only important for some species)?
- (3) To compare the respective efforts of implementing both types of models in a Geographical Information System (GIS).

Study area

The Spring Mountains are located in southern Nevada, 20 km west of Las Vegas, at latitude 36°15' N and longitude 115°45' W (Figure 1). Rising out of the Mojave Desert at 700 m, the Spring Mountains reach a peak elevation of 3600 m within a distance of 10 km, creating a very sharp physiographic gradient. Plant communities range from Mojave Desert shrub at the base, through Joshua tree woodland, sagebrush, pinion/juniper woodland, and a variety of mixed conifers and aspen at the upper elevations. The highest region supports limber and bristlecone pine, with a small area of alpine tundra at the peak. Deep canyons radiate from the highest peak at Mt. Charleston, creating complex mosaics of vegetation across solar radiation gradients on opposing slopes. Natural and anthropogenic fires and disturbances have added to the complexity of the vegetation mosaic. Because the Spring Mountains are isolated by more than 100 km from the nearest mountains with similar elevation, they support many endemic plant and animals at both the species and subspecies level (Nachlinger & Reese 1996).

Most of the Spring Mountains are part of the Toiyabe National Forest (TNF). Because of their proximity to rapidly growing Las Vegas, they have been designated as a National Recreation Area, which will lead to increased visitation and development of recreation facilities. As part of the planning process, TNF contracted with The Nature Conservancy (TNC) to perform an on-the-ground vegetation survey. This survey had basically two complementary objectives: to identify the distribution and abundance of endemic and potentially threatened plants and to create a wide range map of various species and plant communities (Nachlinger & Reese 1996).

The data set

The data set used in this study was sampled by The Nature Conservancy (TNC; Nachlinger & Reese 1996) in the Spring Mountains. This inventory includes all higher plant species. In our study, the data set is split into two subsets. One is used for calibrating the models (*training data set*) whereas the other is used later on to evaluate the quality of model predictions (*evaluation data set*).

Data from 230 plots (generally 20 m × 20 m) were collected in the summer of 1994 (Figure 1). Plots were distributed according to a stratification scheme, using elevation (300 m bands), aspect (four classes, centered on N, S, E and W), slope (0–5°, 5–15° and > 15°), and five geologic substrates. Ocular estimates of percent ground cover for all vascular plant species were collected along with data on soils, litter, and other site-specific characteristics (Nachlinger & Reese 1996). To save access time and maximize the number of samples, plots in each strata were not chosen according to a random scheme, but localized in the field in a more *ad hoc* manner.

For the models presented here, the response variable is the presence/absence of tree and shrub species at each sampled location. Eleven plots containing flowing water with unique riparian vegetation, and four plots supporting unique cliff vegetation (slope > 40°) have been eliminated from the analysis. Therefore, the pared down data set deals with upland vegetation only.

Predictor variables are elevation (*elev*), slope angle (*slo*), northness (*nness*), eastness (*eness*), summer solstice insolation (*ssol*), spring equinox insolation (*esol*) and four topographic position indices calculated at the different smoothing levels: 150 m, 300 m, 1000 m and

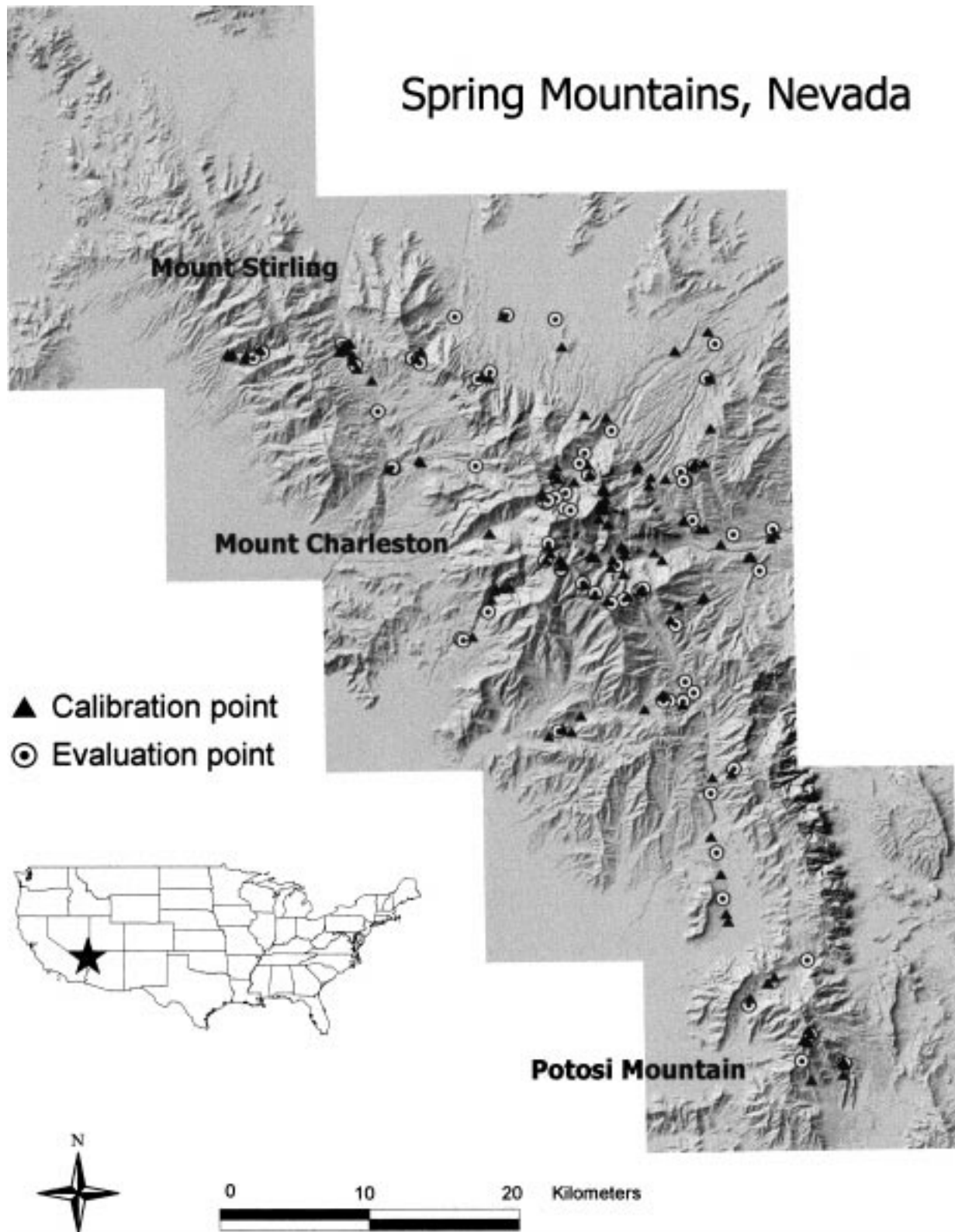


Figure 1. Situation of the study area in the USA and shaded relief of the Spring Mountains, with labeling of the highest elevations (peaks) and the locations of the vegetation sampling sites.

2000 m (*tp150* to *tp2000*). All of the predictors used in this study can be derived entirely from a Digital Elevation Model (DEM; shaded representation in Figure 1) – and hence express terrain related gradients – making the methods widely applicable.

Although we agree that true resource or direct predictors (in the sense of Austin & Smith 1989) would make models more physiological (i.e., snow distribution, solar radiation and wind are the main factor influencing vegetation patterns in a mountainous landscape; Frank 1988), we prefer to use primarily terrain related gradients, because: (1) only environmental predictors that can be mapped across the mountain range in the GIS are included in the analysis; variables such as soil characteristics, local disturbance, and fire history are unfortunately not universally available; (2) spatially-explicit physiologically meaningful predictors tend to be less precise than pure digital topographic characteristics, since they are mostly derived by terrain-sensitive spatial interpolations of climate stations or other punctually-distributed ecologically relevant data (e.g., Hutchinson & Bischof 1993); this is a particular problem in the study area where there is a low density of climate stations; (3) macroclimate is very homogeneous throughout the entire mountain range, so that temperature and precipitation are very closely correlated with elevation; Terrain related gradients are also highly correlated with meso-scale climatic and geomorphological factors that are difficult and imprecise to model directly, such as wind exposure, avalanche and fire frequency, canyon level cold air movements, or soil depth; previous studies showed that terrain related gradients are strong determinants of plant patterns in similar landscapes (e.g., Frank 1988; Burke et al. 1989).

Elevation is sampled from a 30 m resolution DEM covering the study area (Figure 1). All other predictors are then modeled on the basis of the DEM. Slope angle and slope aspect are calculated in the ArcInfo GIS (vers. 7.1; UNIX-based system, ESRI Corp.) by using in-built functions. The circular variable slope aspect is transformed into a continuous north-south gradient (*northness*) and an east-west gradient (*eastness*) by using respectively the sine and cosine transformations. Topographic position indices are calculated at each cell of the DEM by calculating the difference between the elevation of the cell and the mean elevation calculated for all cells of a moving circular window centered on the cell of interest. Four topographic positions are considered, with windows' radii respectively set to 150 m, 300 m, 1000 m and 2000 m. Such indices

are a flexible way to define the relative position of a location along a topographic gradient (ridge top, middle slope or valley). Solar radiation for the equinox and summer solstice was calculated using the Arc Macro Language (AML) program SOLARFLUX (Hetrick et al. 1993), which calculates clear-sky insolation across a DEM accounting for aspect, slope, and topographic shading.

All environmental predictors are finally stored as separate layers in the GIS. Environmental data for each sample point is extracted from the GIS.

One-hundred forty-four plots are randomly chosen from the data set for the training data set, and the remaining 71 are used as the evaluation data set. Models are calculated for trees (all of them; $n = 10$ species) and the most common shrubs ($n = 13$ species) only, as their modeling seems particularly suited at the study resolution (30 m). A finer resolution (10 m or higher) would be preferred for herbaceous species (and particularly for alpine species), although satisfying results were obtained with a similar resolution in the European Alps (25 m; Guisan et al. 1998).

Methods

Statistical methods are used to relate the distribution of plant species (see Nachlinger & Reese 1996, for nomenclatural aspects) to the spatial distribution of environmental predictors. This is done in a 'static' or 'empirical' way by assuming that vegetation is in pseudo-equilibrium with climate and, in turn, vegetation does not influence significantly the regional climate. This contrasts with more dynamic approaches that attempt to model ecological processes (see Korzukhin et al. 1996 for a comparison of both approaches). Several statistical techniques can be used for this purpose. The choice depends primarily upon the type of response variable modeled. A review of some of these techniques can be found in Franklin (1995).

When the response variable is binary (i.e., presence/absence), one can apply for instance a Generalized Linear Model (GLM; see Nicholls 1989), a particular case of multiple regression, with binomial distribution and logistic link (or other adequate links like probit) as used, e.g., by Lenihan (1993).

Another approach is to use Canonical Correspondence Analysis (CCA; Ter Braak 1987) to predict distributions of species or communities (see., e.g., Hill

1991; Gottfried et al. 1998). In this direct gradient analysis technique, main axes of a correspondence analysis (CA; see Hill 1974) are constrained to be a linear combination of environmental descriptors, which makes it very comparable to linear regression.

Other techniques that could be used include simple boolean rules (e.g., Walker et al. 1992), Bayesian approaches (e.g., Breziecki et al. 1993), orthogonalized regressions (Saetersdal & Birks 1997), generalized additive modeling (GAM; e.g., Yee & Mitchell 1991), classification techniques (e.g., Walker & Moore 1988), discriminant analysis (e.g., Frank 1988), neural networks (e.g., Fitzgerald & Lees 1992) or environmental envelopes (e.g., Shao & Halpin 1995). Quantitative model comparison between environmental envelopes (BIOCLIM and HABITAT) and classification (CART) can be found in Walker & Cocks (1991) or between environmental envelopes (BIOCLIM), classification (CART) and a particular bayesian approach (nonparametric classifier) in Skidmore et al. (1996).

Here, we focus on two broadly used techniques, GLM and CCA, as (1) they are relatively straightforward to implement in a Geographical Information System; (2) they represent two very distinct approaches in the sense that CCA allows for the simultaneous modeling of all species included in the study whereas, with GLM, a distinct modeling effort is needed for every species; (3) yet no quantitative comparison have been done between them.

GLM-based modeling

Because the response is binary rather than following a true binomial distribution, generalized linear models (GLM; see Nicholls 1989) are calibrated in S-PLUS (MathSoft Inc.) by specifying a *quasi* distribution with a variance function equal to $\mu(1-\mu)$ and a logistic link function. This alternative is preferred to a binomial distribution with logistic link, as the scale parameter can in this way be automatically estimated rather than set to one (the default for Binomial models). However, with such a model, the estimated coefficients and adjusted values appear to be very similar to those provided by a Binomial GLM. The fit of the model (D^2) is calculated as the proportion of explained deviance (i.e., similar measure as the variance; see McCullagh & Nelder 1983). It can be corrected by the number of degrees of freedom (df) used to fit the model, to provide an adjusted- D^2 (similar to an adjusted- R^2 ; see Weisberg 1980), defined as

$$\text{adj} - D^2 = 1 - [(n - 1)/(n - p)] \cdot [1 - D^2],$$

where n is the total number of observations, p is the number of parameters used to fit the model (see Table 1) and $(n - p)$ is the residual degree of freedom. In order to make easier the final ecological discussion of the models (e.g., interaction terms), we do not orthogonalize the predictors (e.g. through principal component analysis) prior to the model calibration.

GLM models are readily implemented in a geographical information system (GIS). Each model is implemented by building a single formula where each coefficient multiplies its related predictor variable. The results of the calculations are obtained to the scale of the linear predictor (LP) so that the inverse logistic transformation

$$p(y) = \exp(LP)/(1 + \exp(LP)),$$

is then necessary to obtain probability values between 0 and 1 at every cell of the GIS grid. We use a custom function in S-Plus that automatically fits the final model of any species – starting from a pre-defined species specific formula – and provide the D^2 , calibration and evaluation Kappa values as outputs. In a last step, the function automatically writes a species' specific GIS macro (Arcinfo Macro Language; AML) that allows the calculation, from the fitted regression coefficients and for each grid cell in the GIS, of the probability of species occurrence.

CCA-based modeling

CCA-based models were calibrated in CANOCO (ver. 3.12; Ter Braak 1988). Model calibration is very close to linear regression, except that the goodness-of-fit criterion is here to 'minimize the ratio of the mean within-species sum-of-squares of the variate to the overall sum of squares' (Hill 1991). Selection of variables can also be made in a stepwise approach, but it is not species specific (selection is made of the environmental variables explaining successively the highest proportion of variance in the species data as a whole). Axes can be tested for significance through successive Monte Carlo permutations. CCA is appropriate to deal with species data sets that contain many zeros (i.e., absences), but rely on several postulates (e.g., unimodal symmetrical species' response curves, species have equal ecological amplitude and equal maxima)

Table 1. Comparison of predictions from GLM and CCA models. Species' models are listed by decreasing accuracy of the GLM predictions (evaluation Kappa). part. = partially-; Dist. = Disturbance; Aval. = Avalanche.

Species	Ecological requirement	Calibration						Evaluation		
		Presence (out of 144)	No. of predictors	Proportion of explained variance		Kappa		Presence (out of 71)	Kappa	
				GLM	CCA	GLM	CCA		GLM	CCA
<i>Coleogyne ramosissima</i>	—	34	3	0.90	0.47	0.94	0.84	17	0.82	0.76
<i>Yucca brevifolia</i>	—	17	4	0.71	0.27	0.85	0.62	9	0.80	0.54
<i>Pinus longaeva</i>	—	28	5	0.75	0.35	0.88	0.64	13	0.73	0.57
<i>Yucca baccata</i>	—	44	5	0.49	0.37	0.68	0.59	27	0.79	0.80
<i>Ephedra viridis</i>	—	66	4	0.52	0.42	0.72	0.53	37	0.77	0.59
<i>Pinus flexilis</i>	—	15	5	0.66	0.27	0.75	0.44	10	0.72	0.39
<i>Pinus ponderosa</i>	—	25	3	0.60	0.27	0.78	0.56	9	0.70	0.49
<i>Cercocarpus ledifolius</i>	—	46	7	0.56	0.17	0.73	0.61	16	0.69	0.56
<i>Pinus monophylla</i>	—	74	4	0.45	0.09	0.65	0.49	37	0.63	0.55
<i>Artemisia tridentata</i>	—	64	6	0.31	0.02	0.54	0.33	35	0.63	0.42
<i>Juniperus osteosperma</i>	—	61	4	0.47	0.28	0.68	0.61	36	0.61	0.61
<i>Gutierrezia sarosensis</i>	part. Dist.	24	9	0.39	0.16	0.55	0.42	14	0.61	0.42
<i>Abies concolor</i>	—	35	6	0.80	0.40	0.89	0.62	11	0.48	0.40
<i>Ribes cereum</i>	part. Fire	28	10	0.55	0.29	0.71	0.56	17	0.41	0.58
<i>Arctostaphylos pungens</i>	Fire	21	7	0.47	0.08	0.56	0.29	7	0.35	0.20
<i>Quercus gambellii</i>	Fire	22	5	0.58	0.14	0.71	0.62	5	0.31	0.33
<i>Populus tremuloides</i>	Aval./Fire	7	4	0.42	0.10	0.44	0.22	3	0.31	0.25
<i>Gareya flavescens</i>	Fire	22	9	0.52	0.22	0.70	0.52	9	0.28	0.33
<i>Chrysothamnus viscidiflorus</i>	Dist.	20	5	0.17	0.08	0.37	0.15	14	0.22	0.19
<i>Chrysoth. nauseosus</i>	Dist.	11	4	0.23	0.03	0.41	0.11	6	0.21	0.10
<i>Amelanchier utahensis</i>	Fire	16	8	0.46	0.03	0.63	0.31	9	0.13	0.33
<i>Acer glabrum</i>	—	4	2	0.50	0.15	0.66	0.20	2	0	0.18

that may theoretically invalidate its use in some situations. However, Ter Braak (1987) argues that his method is still robust when such postulates are violated, and simulation studies show that CCA is robust in the face of skew and noise (Palmer 1993). It has also been used successfully to predict spatial distribution of binary data (Hill 1991).

An overall measure of the CCA fit is given both by the trace (or total inertia) of the underlying correspondence analysis (CA) and by the proportion of variance in the species' data that is explained by each canonical axis. The trace is the total variance in the species data (i.e., the sum of all eigenvalues). It is measured by the chi-square of the sample-by-species table (Greenacre 1984) divided by N , the table's grand total (see Ter Braak & Smilauer 1998).

The fit of a particular species by k CCA axes is given cumulatively and expressed as a fraction of the

variance of a species. The species variance is calculated as the chi-square of the sample-by-species table divided by species' column total (for more details, see Greenacre 1984, or Ter Braak & Smilauer 1998). The reported fits are the regression sums of squares of the weighted regression of the data for the species, expressed as a fraction of the total sum of squares for the species (i.e. in a similar way as R^2 in GLMs), on the 1 to k ordination axes. The overall percentage of explained variance ('% EXPL' in CANOCO outputs) is obtained by summing all axes. These measures of the fit are discussed in more details in Ter Braak & Smilauer (1998).

In addition, the species-environment correlation can be measured for each axis as the correlation of the respective multidimensional coordinates of the species occurrences in both the species and the environmental space. The later is resulting from multiple regres-

sion predictions of the species coordinates on the environmental variables. A high species-environment correlation does not necessarily mean that an appreciable amount of the species data is explained by the environmental variables (Ter Braak 1988), and thus is not a good measure of the fit.

Calculating the main axes from a linear combination of the original environmental predictors (using the canonical coefficients) readily implements CCA predictions in a GIS. Each axis constitutes a grid layer in the GIS. In fact, it is a map of a key environmental gradient, and is a useful product in its own right. All the axis layers together define the new canonical space. This allows determining the canonical coordinates of each grid cell on the map and calculating its Euclidean distance to each species centroid in canonical space. Distances can be grouped by classes of standard deviation unit and mapped to draw the potential distribution of each species. The whole procedure is carried out by a custom macro (CANOGEN AML) that uses elements of the CANOCO output file as input and produces a potential map for each species included in the data set.

Comparing the two approaches

Although both approaches rely primarily on two different methodologies and mathematical algorithms, they also differ seriously in the way the predictor and the response variables are handled. Whereas the CCA-based approach integrates all predictor and response variables at the same time, in GLMs, a species specific subset of predictors has generally to be selected and each species is modeled in a separate run. CCA models use the variance associated with all predictors, which is best summarized in orthogonal axes, to fit the model of any species. Hence, the linear combination of predictors is exactly the same for all species. In GLMs, only the subset of predictors that allows for the highest deviance reduction is kept in the final model, so that each species distribution is predicted by a possibly different linear combination of predictors. Moreover, while interaction terms can easily be incorporated into GLM and CCA models, they need to be interpreted cautiously.

In CCA, the problem of multicollinearity between predictors is solved automatically by calculating orthogonal axes that minimize the variance. Whereas in GLMs, orthogonalizing the predictors is an alternative that can be made prior to the model calibration (except in the case of partial least square procedures where it

is included in the model calibration; see Heikkinen 1996). However, this procedure is not always necessary if the predictors retained in the final model show little correlation, as was the case in our study.

Finally, in GLMs the predicted values are probabilities, whereas in CCA, a distance to the centroid of species' presence occurrences is calculated in term of standard deviation units. Both types of prediction can be cut back to the presence/absence scale of the response variable by calibrating a threshold value that provides optimal predictions (Guisan et al. 1998).

The best measure of agreement between observed and predicted presence/absence is Kappa (Cohen 1960; see also Monserud & Leemans 1992). Its calculation is based on marginal probability of a contingency table, which is:

$$\kappa = (\theta_1 - \theta_2) / (1 - \theta_2),$$

where $\theta_1 = \sum_i p_{ii}$ and $\theta_2 = \sum_i p_{i+} p_{+i}$. θ_1 is the sum of the diagonal elements p_{ii} in the table, i.e., the agreement between observed and predicted values. Hence, it represents the overall proportion of observed agreement. θ_2 calculates the overall proportion of chance-expected agreement that occurs if the rows are independent of the columns (Monserud & Leemans 1992), and serves to correct θ_1 . It is based on multiplying marginal frequencies of the row and column – p_{i+} and p_{+i} – and summing up each element of the table. Kappa is used as the main measure, in this study, to compare the predictive capacity of both procedures.

Results

GLM modeling

The results of GLM models are given in Table 1. The combination of predictors retained in each model is given in Table 2. GLM models explain a proportion of deviance ($\text{adj-}D^2$) that range from 0.15 (weak fit) to 0.90 (high fit; Table 1). These values are already weighted by $n-p$ the residual number of degree of freedom, with p the number of predictors used in the model (4th column in Table 1) and n the total number of observations (out of a maximum of 144 in the calibration data set).

Models in Table 1 are ordered as to the predictive quality (evaluation Kappa) of their GLM model. Following this classification, four categories of model

Table 2. Detailed formulas of the GLM models. All the terms were significantly retained in the model at the 0.05 confidence level by both a t-test and a Chi-square test of deviance reduction. The function poly stands for polynomials. For instance, poly(slo,2) means a second order of the type $slo + slo^2$. Finally, the double dots (:) between two variable define an interaction term; elev = elevation, slo = slope, tp = topographic position (following number is the size of the moving windows, see text for explanation), eness = eastness, nness = northness, esol = equinox insolation, ssol = summer solstice insolation.

Species	GLM formulas
<i>Abies concolor</i>	poly(elev,2) + poly(nness,2) + tp2000 + nness:tp2000
<i>Acer glabrum</i>	esol ² + tp1000
<i>Amelanchier utahensis</i>	poly(elev,2) + esol + tp300 ³ + nness ² + poly(elev,2):esol + elev:nness ²
<i>Arctostaphylos pungens</i>	poly(elev,2) + nness + esol + nness:esol + poly(elev,2):esol
<i>Artemisia tridentata</i>	poly(elev,2) + tp2000 + eness + poly(elev,2):tp2000
<i>Cercocarpus ledifolius</i>	poly(elev,2) + esol + poly(tp2000,2) + elev:esol + elev ² :tp2000
<i>Chysothamnus nauseosus</i>	elev + nness + ssol + ssol:nness
<i>Chysothamnus visciflorus</i>	elev + slo + nness + ssol + slo:nness
<i>Coleogyne ramosissima</i>	elev + poly(tp300,2)
<i>Ephedra viridis</i>	elev + tp200 + nness + ssol
<i>Gareya flavescens</i>	poly(elev,2) + tp1000 ² + nness + ssol + esol + ssol:esol + ssol:elev + tp1000 ² :nness
<i>Gutierrezia saroscens</i>	poly(elev,2) + nness + ssol + poly(tp2000,2) + elev:nness + ssol:nness + nness:elev ²
<i>Juniperus osteosperma</i>	poly(elev,2) + tp2000 + eness
<i>Pinus flexilis</i>	poly(elev,3) + esol + esol:elev ²
<i>Pinus longaeva</i>	poly(elev,2) + esol + tp2000 + esol:tp2000
<i>Pinus monophylla</i>	poly(elev,2) + tp1000 + elev:tp1000
<i>Pinus ponderosa</i>	poly(elev,2) + tp2000
<i>Populus tremuloides</i>	poly(elev,2) + tp2000 + elev:tp2000
<i>Quercus gambellii</i>	elev ² + elev ³ + tp300 + esol + elev ² :esol
<i>Ribes cereum</i>	poly(elev,2) + tp2000 + esol + ssol + poly(elev,2):esol + ssol:tp2000 + poly(elev,2):tp2000
<i>Yucca baccata</i>	poly(elev,3) + slo + ssol
<i>Yucca brevifolia</i>	elev + elev ³ + tp2000 + nness

are distinguished. Thicker plain lines separate the four model categories in Table 1.

- (1) Models showing a very high Kappa at evaluation (>0.8).
- (2) Models still having a very satisfying value of Kappa ($0.6 < \text{Kappa} < 0.8$).
- (3) Models having a non-satisfying value of Kappa (i.e., <0.5), although enough observations of presence are available to fit a model.
- (4) Models having a Kappa value of zero, because too few observations are available, to both calibrate and evaluate the model.

Interestingly, our results tend to indicate that, for a given species, a very low amount of presence amongst the total observations N may still provide an apparently satisfying fit for the model, although evaluation is actually very poor (Figures 4f and 4h). For example, this is the case with the models for *Amelanchier utahensis* and *Acer glabrum*. Models for some species, i.e. *Yucca baccata* or *Ephedra viridis*, have a low

D^2 (around 0.5) although their evaluation is still considered good on the Kappa scale of agreement (see Monserud & Leemans 1992). Conversely, models for species with high D^2 and high calibration Kappa may have poor evaluation Kappa (*Abies concolor*, in particular). The overall correlation between D^2 and calibration and evaluation Kappa were 0.95 and 0.49, respectively (Figures 4d and 4f). Calibration and evaluation Kappa were only moderately correlated ($r = 0.56$; Figure 4h). All three correlation values were significant at the 97% level (or higher; all p -values <0.03). These relatively low correlation values suggest that the GLM tend to overfit the calibration data set, leading to a poor evaluation fit. Hence, it is suggested that the classical D^2 measure of the fit may not be fully reliable in the case of GLMs, and one should rely more on empirical evaluation of the models.

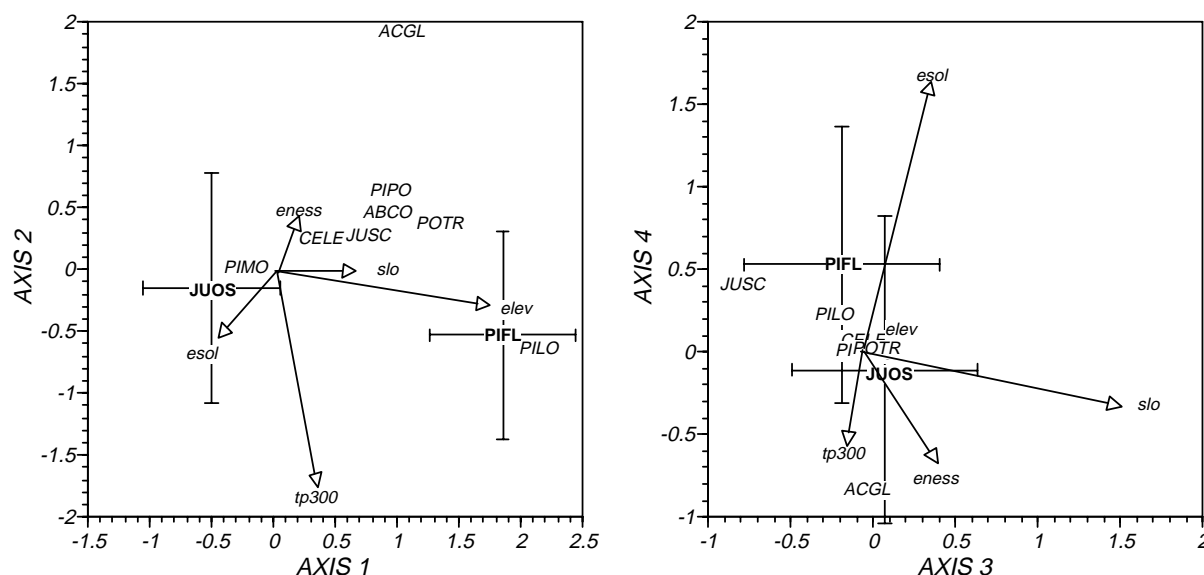


Figure 2. CCA biplots of trees. Species scores are linear combination scores. Species codes are the first two letters of the genus and first two letters of the species names (see Table 2 for full names). Tolerances (± 1 s.d.) are shown for *Juniperus osteosperma* (JUOS) and *Pinus flexilis* (PIFL). (a) Axis 1 versus axis 2. (b) Axis 3 versus axis 4.

CCA-based modeling

The overall ordination used five variables that were selected using the forward selection procedure in CANOCO. Scaling 2 (species scores are weighted average site scores) was used because the sites covered virtually the entire elevation range of the mountains, and most species distributions fell well within the range of sites. The ordination biplots for the 10 tree species shows the overall results of the CCA ordination (Figure 2). The variables retained are: elevation (*elev*), slope (*slo*), equinox insolation (*esol*), topographic position 300 (*tp300*) and eastness (*enness*). The eigenvalues, and several measures of the fit of the 4 axes are shown in Table 3. Cumulative proportions of variance of the species data are quite low (Table 3), but this is usual, as presence-absence data are often very noisy (Ter Braak 1988). Moreover, an ordination diagram that explains only a low percentage of variance may still be quite informative (Gauch 1982). The evaluation Kappas tend to be better than one might expect from the low proportion of variance explained, indicating that the spatial context is being relatively well modeled. Canonical coefficients, t-values, and correlation of environmental variables with ordination axes are given in Table 4.

Axis 1 (eigenvalue = 0.71) is dominated by elevation (Figure 2a, Table 4), and is interpreted as the temperature-precipitation gradient as one proceeds

Table 3. Results from the CCA analysis. Characteristics of the four first ordination axes. FR = fraction, SPEC: species data, ENV: set of environmental predictors.

	AX1	AX2	AX3	AX4
Eigenvalue	0.71	0.14	0.09	0.06
FR explained	0.68	0.13	0.08	0.06
FR extracted	0.23	0.09	0.05	0.05
R (spec, env)	0.95	0.65	0.59	0.48
FR fitted	0.11	0.02	0.01	0.01

from desert to subalpine forest. The tree species sort out well along the elevation gradient – *Juniperus osteosperma* is the first tree encountered as one ascends the mountain range, closely followed by *Pinus monophylla*. These two species overlap considerably and form the common pinyon-juniper woodland recognized across much of the southwest United States. *Cercocarpus ledifolius*, *Pinus ponderosa*, and *Abies concolor* are dominant at middle elevations, *Populus tremuloides* is a relatively high elevation species, and high elevation forests consist of *Pinus flexilis* and *Pinus longeva*, with the latter forming pure stands at the highest elevations. This is the widely recognized elevational order of trees in Great Basin mountain ranges (Lanner 1983). Elevation is modified by equinox insolation (*esol*), so that effective elevation is higher in low

insolation sites, and lower in high insolation sites. The modifying effect of insolation can be readily seen in the field when opposing south-facing and north-facing slopes of canyons support lower and higher elevation tree species, respectively.

The shrubs also sort out along Axis 1. The cluster of species with low Axis 1 scores are typical high desert shrubs, those around the origin are mid-elevation montane chaparral species, and *Ribes cereum* is a high elevation shrub associated with pine-fir forests.

Axis 2 (eigenvalue 0.14) is dominated by topographic position (*tp300*), which is interpreted as a ridge top to valley bottom drainage gradient, as well as a broad surrogate for micro-climate (wind exposure, cold air drainage), soil depth and frequency of avalanche disturbance. The major effect of Axis 2 is to differentiate *Acer glabrum*, a tree of moist ravines. The relatively lower scores of *Pinus flexilis* and *Pinus longeava* on Axis 2 indicate that these species are more often found on ridge tops than in canyons, but that may also be a function of their presence at the highest elevations where ridges are common and canyons are rare. Shrubs on Axis 2 show relatively little differentiation, but *Quercus gembellii* exhibits a slight preference for canyons, and *Yucca brevifolia* has a preference for ridges.

Axis 3 and Axis 4 are relatively minor (eigenvalues <0.10), but still significant in the Monte Carlo test. Most of the trees are grouped close to the origin. The low Axis 3 score for *Juniperus scopulatum* indicates a preference for flatter areas, and the relatively high score of *Pinus flexilis* on Axis 4 indicates a slight preference for warmer slopes receiving more solar radiation. For the shrubs, most species cluster near the origin, but *Garreya flavescens* exhibits a preference for steeper and warmer slopes receiving more solar radiation.

The tolerances (1 standard deviation; s.d.) for *Juniperus osteosperma* and *Pinus flexilis* are shown to illustrate that tolerances are an important component of CCA that are rarely presented, and form a key to the CANOGEN projections shown below. Species tolerances along Axis 1 for trees ranged from 0.39 to 0.64 s.d. units, for shrubs they ranged from 0.31 to 0.83. Tolerances were larger on Axis 2 (0.82 to 1.56 for trees, 0.68 to 1.29 for shrubs). While CCA attempts to equalize species tolerances, it is not possible in complex data sets, and the different lengths of the tolerances, reflecting in part the niche breadth, are

an important part of the geographic projections shown below.

Overall the CCA model explained 15% of the variance in species presence-absence. Of course this varies from species to species (Table 1). Maximum fit was 47% for *Coleogyne ramosissima*, and 10 out of 22 species had greater than 25% of the variance explained. At the other end of the spectrum, however, several species had less than 5% of the variance explained (*Artemisia tridentata*, *Arctostaphylos pungens*, *Pinus monophylla*).

The best evaluation Kappa was 0.80 for *Yucca baccata*, and eight other species had acceptable evaluation Kappas from 0.50 to 0.79, and the remainder had poor evaluation Kappas ranging down to 0.11 for *Chrysothamnus nauseosus*. There is not necessarily a high correspondence between % variance explained and evaluation Kappa – *Pinus monophylla* had only 9% of its variance explained by the calibration data set, but had an evaluation Kappa of 0.55. *Garreya flavescens* had 22% of its variance explained in the CCA model but evaluation Kappa was poor at 0.33. Overall, however, there is a strong correlation between % variance explained and both calibration and evaluation Kappa ($r = 0.78$ and 0.72 respectively; Figure 4e and 4g). Calibration and evaluation Kappa were strongly correlated (Figure 4i; $r = 0.84$). All three correlation values were significant at the 99% confidence level (all p -values < 0.01).

Comparing the two approaches

The predictor variables for each GLM model were chosen from the overall set used in the CCA models. The major contrast between the two approaches is that CCA uses the same predictor variables for all species, while GLM chooses species-specific sets of predictors. The treatment of elevation – the single most important determinant of species distributions in this mountain range – in GLM illustrates some important differences. Elevation shows up as quadratic function in 13/22 species models (all mid-to high-elevation species), as a linear term in 4/22 models (all low elevation shrubs), and as a third degree polynomial in 4 models indicating a skewed distribution (Table 2). Topographic position at various scales (*tp300*, *tp1000*, or *tp2000*) shows up in 17 models. Equinox insolation, summer solstice insolation, or northness show up in 17 models, indicating that solar radiation loads are important local predictors of species distributions. GLM

Table 4. Results from the CCA analysis. Relationship of axes with each environmental predictor. In turn, each axis can be defined as a combination of the environmental predictors; t = t -value of regression coefficient; r = correlation coefficient.

Variable	AX1	AX2	AX3	AX4	t-AX1	t-AX2	t-AX3	t-AX4	r-AX1	r-AX2	r-AX3	r-AX4
elev	1.04	0.10	-0.44	0.44	29.04	0.83	-3.07	2.33	0.93	-0.11	0.04	0.03
slo	-0.08	-0.09	1.12	-0.19	-2.28	-0.78	8.12	-1.06	0.43	0.00	0.49	-0.08
esol	-0.15	-0.31	0.36	0.82	-4.82	-2.97	3.01	5.10	-0.27	-0.22	0.10	0.40
tp300	-0.13	-0.94	0.00	-0.46	-4.22	-8.91	-0.02	-2.77	0.16	-0.60	-0.06	-0.15
eness	0.00	0.15	0.24	-0.29	0.12	1.51	2.05	-1.87	0.09	0.15	0.11	-0.19

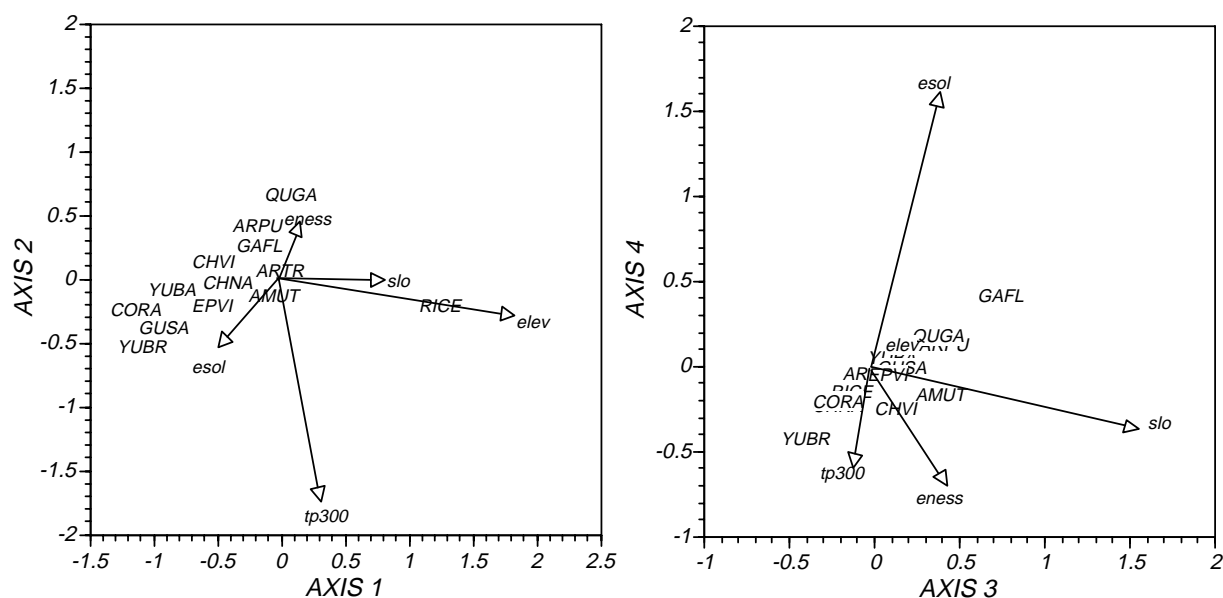


Figure 3. CCA biplots of shrubs. Species scores are linear combination scores. Species codes are the first two letters of the genus and first two letters of the species names (see Table 2 for full names). (a) Axis 1 versus axis 2. (b) Axis 3 versus axis 4.

models often have complex interaction terms (15/22 models), many of which are hard to interpret.

Overall, GLM models explain a much higher proportion of variance than CCA-based models (in most cases; see Figure 4a). GLMs also show in most cases a better agreement between observed and predicted presence/absence at the evaluation (Figure 4c). In a few cases – for *Ribes cereum*, *Amelanchier utahensis* and *Acer glabrum* – CCA models have a much better evaluation Kappa. In other cases, they are similar (max difference = 0.05), as in the case of the models for *Yucca baccata*, *Juniperus osteosperma*, *Quercus gambellii*, *Garreya flavescens*, *Chrysothamnus viscidiflorus* and *Ch. nauseosus*. In all other cases, GLMs have a better evaluation Kappa.

The rankings of measures of fit among species were similar for both techniques. Generally, a species

whose distribution is well modeled by GLM (relative to the others GLM models) is also well modeled by CCA (relative to the other CCA models). The correlation among various measures of the fit between the two methods are strong ($r = 0.76, 0.87$, and 0.82 for % var/dev explained, calibration Kappa, and evaluation Kappa respectively; Figures 4a–c; highly significant, with p -values < 0.001). In both cases, species for which model predictions have the poorest accuracy are either disturbance or fire related (category 3 in Table 1), or species for which too few observations were available to calibrate and evaluate the model (category 4 in Table 1).

Potential distribution maps

Once the coefficients are estimated, the models can be projected into geographic space. Two examples,

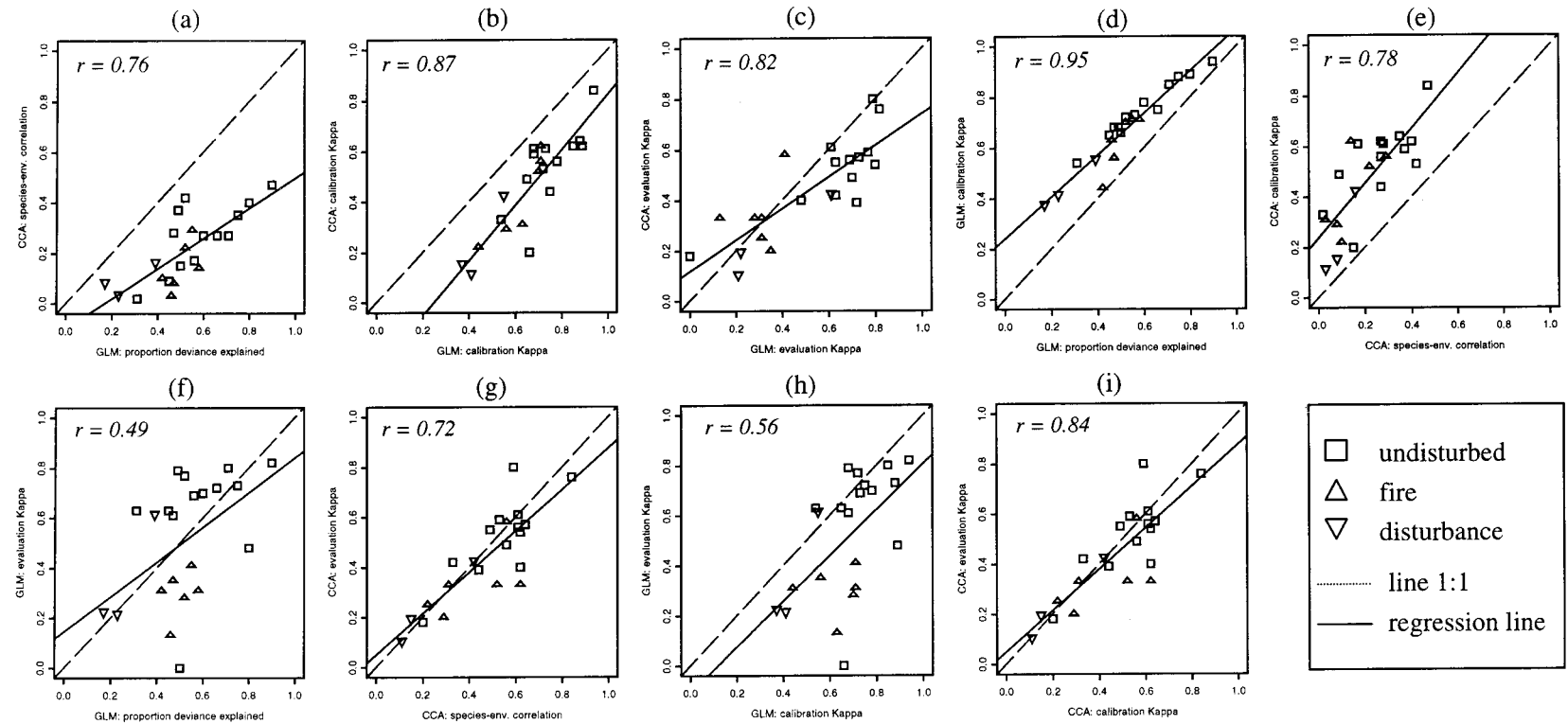


Figure 4. Comparing GLM versus CCA models (a-c) and, within each approach, calibration versus evaluation (d-i). (a) Proportion of explained variance versus species-envir. correlation. (b) Calibration Kappa. (c) Evaluation Kappa. (d) GLM: fit versus calib. Kappa. (e) CCA: species-env. corr. versus calib. Kappa. (f) GLM: fit versus eval. Kappa (g) CCA: species-env. corr. versus eval. Kappa (h) GLM: calib. versus eval. Kappa. (i) CCA: calib. versus eval. Kappa. Each point represents a single species, symbolized according to its disturbance context.

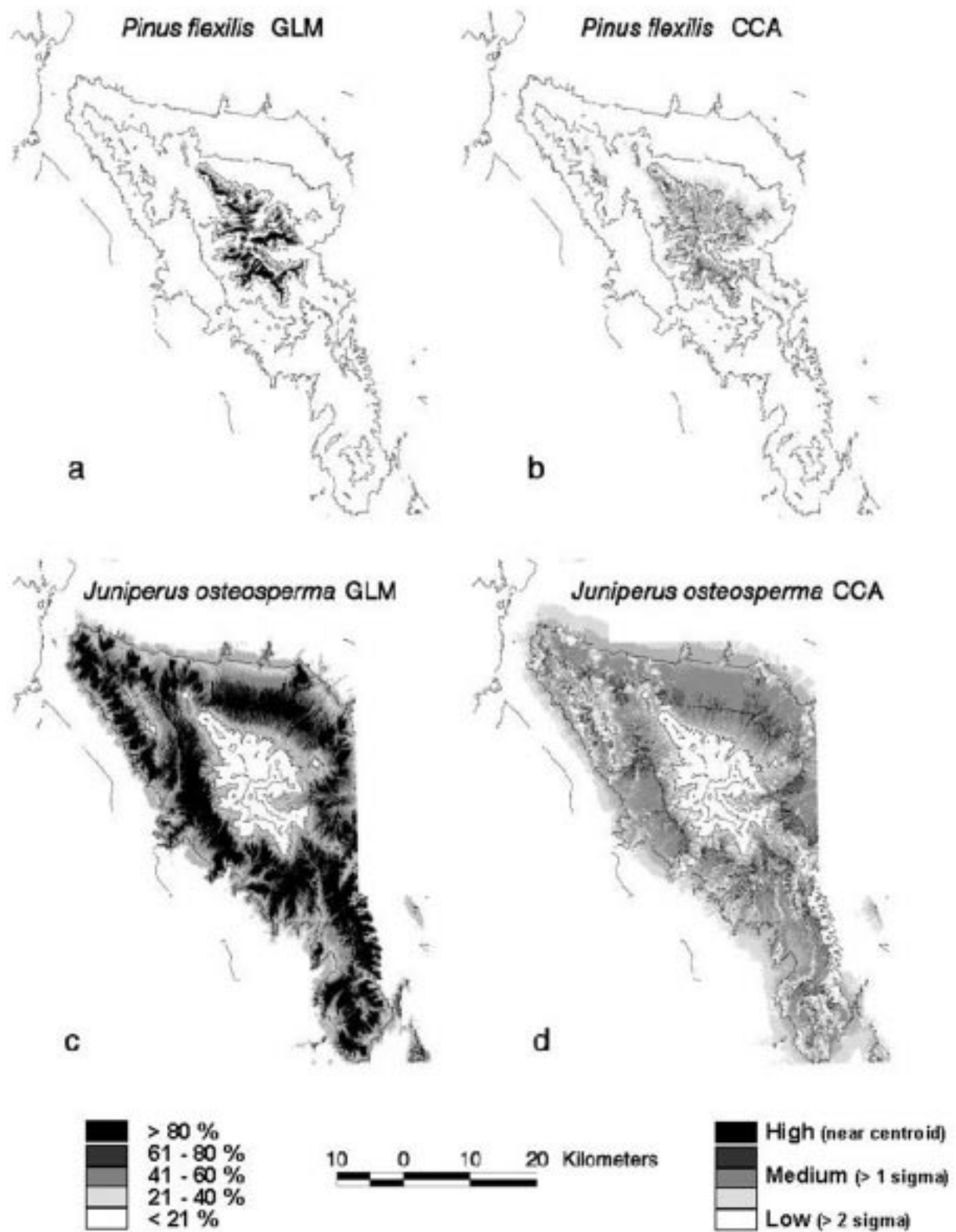


Figure 5. Potential distribution map of *Pinus flexilis* in the Spring Mountains (Nevada), (a) modeled by GLM, and (b) modeled by CCA. Potential distribution map of *Juniperus osteosperma* in the Spring Mountains (Nevada), (c) modeled by GLM, and (d) modeled by CCA. GLM predictions are probabilities (given in percent in the legend). CCA predictions are in S.D. unit distance from the centroid of the species observation in the ordination space.

for *Pinus flexilis* and *Juniperus osteosperma*, are respectively given in Figures 4a–d. On a broad scale, predicted distributions from both approaches look similar, which is largely a function of the dominance of elevation.

The probability of species occurrence is low at the lower elevation limit, increases to a maximum, and then decreases towards the upper limit. Both the inclusion of quadratic elevation terms in GLM models as well as the reciprocal averaging algorithm in CCA (that consider bell-shaped unimodal responses) effectively capture this feature of the distributions. Many of the finer scale details of distribution are related to the effects of insolation and topographic position.

The *Pinus flexilis* model in CCA produces a much broader predicted range than the GLM model. The CCA model predicts that *Pinus flexilis* occurs on Mt. Stirling (the area NW of the main distribution) and on Potosi Mountain (the small area to the south), while the GLM model predicts a low probability of occurrence in a small area on Mt. Stirling only. Small stands of *Pinus flexilis* do indeed occur on Mt. Stirling, but the species is not known from Potosi. Both models indicate that the species is more likely to occur on south-facing slopes within the overall elevation range. While CCA may provide a better approximation of the overall species range, it was notably poorer than GLM in terms of both calibration Kappa (0.44 in CCA, 0.75 in GLM) and evaluation Kappa (0.39 in CCA, 0.72 in GLM) in Table 1.

The projected distributions of *Juniperus osteosperma* are very similar in extent, and the fine scale differences are related to the fact that the GLM model does not include insolation as a predictor, which limits the spatial variation due to opposing slopes. Calibration and evaluation Kappa were similar for both models (Table 1).

Discussion

In most cases, GLM models clearly explain a higher proportion of variance (i.e., deviance) than CCA models. This might result from the fact that GLMs, through a stepwise selection of predictors, allow one to fit more accurately the specific ecological (or realized) niche of a species (see Table 2). However, when only a very few presence are observed for a species amongst the total number of sampled sites (e.g. rare and uncommon species), CCA appears as successful or better than GLM for modeling geographical

distribution. For example, the GLM model for *Acer glabrum* (4 occurrences in the calibration set, 2 occurrences in the evaluation set) does not include elevation, although it is obvious that this species occupies a relatively narrow high elevation range, and is restricted to cool slopes and gullies. This difference can be explained by considering how CCA establishes the species-environment relationships. In CCA, species are not independently related to environmental predictors. On the contrary, the main variance trends in the species' data (i.e., the *latent* gradients) are constrained to be combinations of environmental variables. Hence, as species with similar ecological requirements tend to be similarly distributed, they will have similar scores in environmental – or rather canonical – space. A rare species that is associated in most cases with a frequent species will benefit from the larger number of occurrences of the latter to have its ecological requirements (in terms of CCA scores) better defined. This is presently not possible with GLMs, although it might pave the way for new research directions. In particular, designing systems of simultaneous regression equations (simultaneous GLMs) might allow the improvement of most current plant distribution models by integrating species co-occurrence/exclusion information (Guisan 1997).

From the ecological point of view, both approaches attest the limitation of modeling distribution of species that are strongly related to fire and disturbance history. Fire is very difficult to map, in particular because it is time-dependent. Although a site never affected by fire will not exhibit the same vegetation as a site at least once affected by fire, the proximity in time of the last fire and the recurrence intervals are also important parameters to include in models. However, unless historical fire maps of the study area have been systematically drawn or satellite/aerial photographs are available for many years back (at least 50), such information is rarely available to modelers. Site-specific fire histories can be established by dendrochronology, but such studies are labor and cost intensive, and impossible to establish across broad areas. Other species show dependence on disturbance rather than on fire. Avalanches are common in the Spring Mountains and avalanche paths tend to be dominated by *Populus tremuloides*, a species that has low calibration and evaluation Kappa. Flash floods often occur in desert washes, also, and species such as *Chrysothamnus nauseosus* thrive in disturbed washes. Human disturbances can potentially modify the natural distribution of species. Proximity to roads, settlements

or ski resorts can be used as indices of direct human disturbance, but the scale of much human disturbance is often very fine and below the resolution of the GIS. Other disturbances such as pollution can be more difficult to map and may translate into systematic error components in the predicted distribution of sensitive species. However, the predicted ranges of disturbance related species can indicate areas and physiographic conditions that, over a longer time frame, are more likely to experience fire or other disturbances. For example, the projected range of *Populus tremuloides* may show areas that are susceptible to avalanches and may have had aspens at various times in the past, and potentially in the future.

CCA appears easier than GLMs to implement into Geographical Information Systems (GIS). A drawback of this approach is that predictions are not probabilistic but are expressed as a distance from the centroid of each species. An additional step may be taken here, to run a logistic regression of species against the axes scores (Hill 1991), and use those fitted parameters to generate maps, but then the procedure loses some elegance. With GLMs, predictions are probabilistic but a distinct regression equation needs to be written for each species in the GIS. This makes the implementation of GLM models for many species less straightforward than with CCA, in which the whole procedure can be written as a single macro function generating all species maps at the same time.

CCA also provides a much broader view of general ecological gradients in a study area. Because the final choice of environmental loading on each axis is the result of a weighted averaging algorithm that uses all species, overall ecological gradients (i.e. the temperature/precipitation gradient defined by elevation as modified by insolation) will be made clearer than in any individual GLM model. For community delineation and interpretation, CCA is clearly more valuable because it deals with the co-occurrence of species. It also provides the overall gradient of biodiversity, making it useful for ecosystem management applications. This application of CCA was central in the biodiversity planning for the Spring Mountains National Recreation Area (Nachlinger & Reese 1996).

Conclusion

In a general manner, spatial distribution of individual species is better modeled by Generalized Linear Modeling (GLM) than by Canonical Correspondence

Analysis (CCA). Overall, both approaches provide a similar ranking of model quality. Results show that the less a species is fire- or disturbance-related, the better its potential geographical distribution within the study area can be modeled. The few cases where CCA provides a clearly better evaluation than GLMs are all related to species having a relatively low amount of presence recorded amongst the total number of sampled sites. Hence, as a result, rare species might be better modeled by CCA whereas frequent to moderately frequent species might be better modeled by individual GLMs that better fit the ecological preference of each species.

Acknowledgements

Antoine Guisan was supported by the Swiss National Science Foundation (SNSF). Peter and Helen Bing provided salary funds for S.B. Weiss and A.D. Weiss. Funding for fieldwork in the Spring Mountains was provided by the USDA Forest Service and the US Fish and Wildlife Service. Our best thanks go to Prof. Paul Ehrlich and Dr Carol Bogg at the Center for Conservation Biology, Stanford University, who gave us full support to conduct this research in the GIS laboratory while Dr Guisan was resident as a SNSF funded post-doc. Alicia Torregrossa contributed to the fieldwork and the conceptualization and application of CCA modeling. Environmental Systems Research Institute, Inc. (ESRI) has provided generous donations and support for the CCB GIS Lab. We would also like to thank Jan Nachlinger and The Nature Conservancy (TNC) for providing us the opportunity to collaborate on the Spring Mountains' project. Dr J.-P. Theurillat should also be thanked for letting A. Guisan revise the manuscript within the framework of the PPE/ECOCLINE SNSF project (University of Geneva). Finally, our sincere thanks go to Julie Warrillow who undertook a first linguistic revision of the paper.

References

- Austin, M. P. & Smith T. M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35–47.
- 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 Beta-function. *J. Veg. Sci.* 5: 215–228.
- Brzeziecki, B., Kienast, F. & Wildi, O. 1993. A simulated map of the potential natural forest vegetation of Switzerland. *J. Veg. Sci.* 4: 499–508.

- Burke, I. C., Reiners, W. A., & Olson, R. K. 1989. Topographic control of vegetation in a mountain big sagebrush steppe. *Vegetatio* 84: 77–86.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20: 37–46.
- Ferrer-Castán, D., Calvo, J. F., Esteve-Selma, M. A., Torres-Martínez, A. & Ramirez-Díaz, L. 1995. On the use of three performance measures for fitting species response curves. *J. Veg. Sci.* 6: 57–62.
- Fitzgerald, R. W. & Lees, B.G. 1992. The application of Neural Networks to the floristic classification of remote sensing and GIS data in complex terrain (I). Vol. 3, Pp. 2–10. In: *Proceedings of the 6th Australian Remote Sensing Conference*, Wellington, N.Z., November 1992.
- Frank, T. D. 1988. Mapping Dominant Vegetation Communities in the Colorado Rocky Mountain Front Range with Landsat Thematic mapper and Digital Terrain Data. *Photogr. Eng.* 54: 1727–1734.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geog.* 19: 474–499.
- Gauch, H.G. 1982. *Multivariate Analysis in Community Ecology*. Cambridge University Press, Cambridge.
- Gottfried, M., Pauli, H. & Grabherr, G. 1998. Predictions of vegetation patterns at the limits of plant life: a new view of the alpine-nival ecotone. *Arctic Alpine Res.* 30: 207–221.
- Greenacre, M. J. 1984. *Theory and Applications of Correspondence Analysis*. Academic Press, London.
- Guisan, A. 1997. *Distribution de taxons végétaux dans un environnement alpin: Application de modélisations statistiques dans un système d'information géographique*. Thèse de doctorat présenté à la Faculté des Sciences de l'Université de Genève (PhD Dissertation, University of Geneva). No 2892. 186 pp. + annexes and maps.
- Guisan, A., Theurillat, J.-P. & Kienast, F. 1998. Using static modeling to predict potential distributions of species in an alpine environment. *J. Veg. Sci.* 9: 65–74.
- Heikkinen, R. K. 1996. Predicting patterns of vascular plant species richness with composite variables: a meso-scale study in Finnish Lapland. *Vegetatio* 126: 151–165.
- Hetrick, W. A., Rich, P. M., Barnes, F. J., & Weiss, S. B. 1993. GIS-based solar radiation flux models. Pp. 132–143. In: *American Society for Photogrammetry and Remote Sensing Technical Papers*. Vol. 3, GIS photogrammetry, and modeling.
- Hill, M. O. 1974. Correspondence analysis: a neglected multivariate method. *Appl. Stat.* 23: 340–354.
- Hill, M. O. 1991. Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *J. Biog.* 18: 247–255.
- Huisman, J., Olff, H. & Fresco, L. M. F. 1993. A hierarchical set of models for species response analysis. *J. Veg. Sci.* 4: 37–46.
- Hutchinson, M. F. & Bischof, R. J. 1983. A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to Hunter Valley, New South Wales. *Austral. Meteorol. Mag.* 31: 179–184.
- Jongman, R. H. G., Ter Braak, C. J. F. & van Tongeren, O. F. R. 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge UK.
- Korzukhin, M. D., Ter-Mikaelian, M. & Wagner, R. G. 1996. Process versus empirical models: which approach for forest ecosystem management? *Can. J. For. Res.* 26: 879–887.
- Lanner, R. M. 1983. *Trees of the Great Basin: A Natural History*. University of Nevada Press, Reno, NV.
- Lenihan J. M. 1993. Ecological responses surfaces for north American tree species and their use in forest classification. *J. Veg. Sci.* 4: 667–680.
- McCullagh, P. & Nelder, J. A. 1983. *Generalized Linear Models*. Monographs on Statistics and Applied Probability, Chapman and Hall, London.
- Monserud, R. A. & Leemans, R. 1992. Comparing global vegetation maps with the Kappa statistic. *Ecol. Modelling* 62: 275–293.
- Nachlinger, J. & G. A. Reese. 1996. Plant community classification of the Spring Mountains National Recreation Area, Clark and Nye Counties, Nevada. Unpublished report on file with Toiyabe National Forest, Spring Mountains National Recreation Area, Las Vegas, NV. 104 pp. + appendix.
- Nicholls, A. O. 1989. How to make biological surveys go further with generalized linear model. *Biol. Conserv.* 50: 51–75.
- Oksanen, J. 1997. Why the beta-function cannot be used to estimate skewness of species responses. *J. Veg. Sci.* 8: 147–152.
- Palmer, M. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74: 2215–2230.
- Saetersdal, M. & Birks, H. J. B. 1997. A comparative ecological study of Norwegian mountain plants in relation to possible future climatic change. *J. Biog.* 24: 127–152.
- Shao, G. & Halpin, P. N. 1995. Climatic controls of eastern North American coastal tree and shrub distributions. *J. Biog.* 22: 1083–1089.
- Skidmore, A. K., Gauld, A. & Walker, P. 1996. Classification of Kangaroo habitat distribution using three GIS models. *Int. J. Geographical Information System* 10: 441–454.
- Ter Braak C. J. F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69: 69–77.
- Ter Braak, C. J. F. 1988. CANOCO: an extension of DECORANA to analyze species-environment relationships *Vegetatio* 75: 159–160.
- Ter Braak, C. J. F. & Smilauer, P. 1998. *CANOCO Reference Manual and User's Guide to CANOCO for Windows*. Software for Canonical Community Ordination (version 4). Centre for Biometry Wageningen (Wageningen, NL) and Microcomputer Power, Ithaca NY, USA, 352 pp.
- Walker, P. A. & Moore, D. M. 1988. SIMPLE: an inductive modeling and mapping tool for spatially-oriented data. *Int. J. Geog. Inf. System* 2: 347–363.
- Walker, P. A. & Cocks, D. 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecol. Biog. Letters* 1: 108–118.
- Walker, R. E., Stoms, D. M., Davis, F. W. & van Wagtenonk, J. 1992. Modeling potential natural vegetation from a topographic gradient in the southern Sierra Nevada, California. Pp. 794–803. In: *Proceedings GIS/LIS-92*. Bethesda, MD: ASPRS.
- Weisberg, S. 1980. *Applied Linear Regression*. John Wiley and Sons, New York.
- Yee, T. W. & Mitchell, N. D. 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* 2: 587–602.