

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1994 - 6th Annual Conference Proceedings

PREDICTING COMMON CRUPINA HABITAT WITH GEOGRAPHIC AND REMOTE SENSING DATA

Timothy S. Prather

Bahman Shafii

Robert H. Callihan

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Prather, Timothy S.; Shafii, Bahman; and Callihan, Robert H. (1994). "PREDICTING COMMON CRUPINA HABITAT WITH GEOGRAPHIC AND REMOTE SENSING DATA," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1354>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

PREDICTING COMMON CRUPINA HABITAT WITH GEOGRAPHIC AND REMOTE SENSING DATA

Timothy S. Prather, IPM Specialist, University of California, Kearney Agricultural Center, 9240 S. Riverbend Ave., Parlier, CA 93648; **Bahman Shafii**, Director, Statistical Program, and **Robert H. Callihan**, Professor, Division of Plant Science, College of Agriculture, University of Idaho, Moscow, ID 83844

ABSTRACT

Field surveys for common crupina, as part of an eradication program, are time intensive and could be made more efficient if common crupina habitat could be predicted. Slope, aspect, and vegetation data were used as generalized plant community variables to predict common crupina habitat using a transformed logistic regression. Models were constructed using either aspect or slope as an explanatory variable such that one model predicted the overall effect of either slope or aspect and a set of models predicted the effect of slope or aspect at each of three vegetation classes. A second data set was used to validate the prediction equations for slope and aspect. The proposed models fit the data well and validations were successful as indicated by analysis of residual plots. The probability of finding common crupina was highest for southeast to southwest aspects. In addition, common crupina was most likely to occur, overall, at 25 to 30% slope with decreasing probability at gentler and steeper slopes. Slope models fitted at each vegetation class indicated maximums at 25 to 30% slope for forest and mesic grassland areas but the maximum for arid grasslands was 50% slope. A field detection survey of common crupina that was directed according to probability of occurrence differences along aspect and slope gradients could reduce the area surveyed to 34 to 42%, respectively, of the study area (using a probability cutoff of 30% of the model's maximum). Detection surveys directed according to slope models would find 14% more common crupina than aspect models but would survey 8 to 11% more area. Models that considered vegetation class, when contrasted with models that did not consider vegetation class, did not decrease the total area surveyed while maintaining the same percentage of common crupina found.

Key words: Logistic regression, common crupina, geographic information system, weed eradication

1. INTRODUCTION

Plant invasions increase as our world becomes a global community where travel between countries with different floras becomes commonplace. Eradicating new plant species may maintain distinct floras and prevent further economic hardship to cropped and non-cropped interests. Eradication differs from weed control in that eradication is the destruction of every propagule of a species from an area surrounded by natural or

manmade barriers that prevent reinvasion except by human intervention (Zamora et al. 1989a).

There are three important components to an eradication program: detection, removal and site evaluation, and monitoring. The use of a geographical information system (GIS) to assist in detection, evaluation and monitoring of common crupina (*Crupina vulgaris*) has been discussed (Prather and Callihan 1993) but a predictive model of plant occurrence was not developed. A predictive model of plant occurrence would improve detection survey efficiency by allowing allocation of survey resources to areas that are most likely to contain the species.

A feasibility and strategy study for the eradication of common crupina was conducted and methods developed to assist in eradication (Zamora et al. 1989b). Suggested areas to survey consisted of open areas on southern slopes (Zamora et al. 1989b), resulting in large areas that were surveyed. Plants usually are associated with a specific subset of plant communities found within a region. These plant communities are associated with specific slopes and aspects. Further definition of plant communities can be attained by including general vegetation classes based on reflectance data received by satellite (Landsat). Developing a model based on slope and aspect data would be an indirect measure of the plant communities present within the area of interest. The objective of this research was to develop a predictive model(s) of common crupina occurrence using slope, aspect, and remote sensing data that would identify specific habitats to focus survey efforts.

2. METHODS

Maps containing common crupina locations that had been ground surveyed, were digitized into a Geographic Information System (GIS). Elevation data were purchased from the United States Geologic Survey and a terrain model was built using PC/ARCINFO (Environmental Systems Research Institute Inc., Redlands, CA 92373) software. A Landsat image (July 1989) was imported into IDRISI (Clark University, Graduate School of Geography, Worcester, MA 01610) software and the reflectance values were clustered into 10 general classes. Dominant plant species at 50 sites were recorded to characterize the plant communities within the ten general classes. These data were transferred into PC/ARCINFO after classification.

Each layer of geographic data was then overlaid and a new geographic data base was created that included polygons containing information on presence or absence of common crupina, slope, aspect and vegetation class. Slope is defined as the change in elevation over a given horizontal distance, i.e., a 1 m elevation change in 100 m is a 1% slope. Aspect is the departure, measured in degrees, from true north and increases numerically in a clockwise fashion, i.e., due east is 90° and due south is 180°. The completed data base contained 33,000 data points representing an area of 2,200 km².

Roughly half of the data were used for development of models and the other half were used for model validation.

The explanatory variables were slope, aspect and vegetation classes. The slope variable was divided into 25 classes (Table 1). There tends to be greater diversity of plant communities at shallower slope within the canyons of north-central Idaho. Slope classes were narrowed at shallow slopes to account for community diversity. The 25th class consisted of vertical slopes and errors in the terrain model located at the edge of individual 7.5" maps. The 25th class was deleted from the data set, leaving 24 classes for the analysis. The aspect variable was developed with consideration that the greatest change in plant communities in the canyonlands of north central Idaho is in a north-south gradient. There were 18 aspect classes developed in 20° increments with 10° increments on each east or west direction from the north-south gradient (Table 2). Vegetation classes were collapsed into three classes representing forested areas, mesic grassland areas, and arid grassland areas.

Assigning probabilities of finding common crupina at specific levels of slope and aspect values would aid in the decision making process for detection surveys. However, in this case, the response variable was binary requiring a proper transformation. A simple approach to the binary response regression is the use of the logistic function, given by:

$$E(y_i) = p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})}}, \quad i = 1 \dots m, \quad (1)$$

where p_i is the proportion of data points at each level of x_i (i = levels of x and k = number of independent variables) where $y_i = 1$ (common crupina present). Hence, the logit transformation is given by:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}. \quad (2)$$

Since there were repeated observations at each level of x_j , the sample proportion of 1's were used to obtain estimates of p_j as:

$$\tilde{p}_i = \frac{r_i}{n_i}, \quad i = 1, 2 \dots m \quad (3)$$

where r_j is the number of observations containing crupina at each x_j and n_j = the total number of observations taken at x_j . Imposing the estimate in (3) or (2), one would obtain the transformed logistic regression model:

$$\ln\left(\frac{\tilde{p}_i}{1-\tilde{p}_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, m, \quad (4)$$

whose parameters can now be estimated using ordinary least squares. However, since p_i varies with the levels of regressor variables, the error variance is not homogeneous and thus weighted least squares should be considered. In this case, at each fixed level of x_j the variance of the response is given by:

$$\text{var} \ln\left(\frac{\tilde{p}_i}{1-\tilde{p}_i}\right) \approx \frac{1}{n_i \tilde{p}_i (1-\tilde{p}_i)}, \quad (5)$$

therefore weights, which are the reciprocals of variances at each level of x_j , are given by:

$$w_i = n_i \tilde{p}_i (1-\tilde{p}_i), \quad i = 1, 2, \dots, m. \quad (6)$$

The appropriate estimator is the generalized least squares estimator given by:

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \quad (7)$$

where

$$\mathbf{V} = \text{diag}[\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2]. \quad (8)$$

The weight matrix in this case is given by:

$$= \text{diag}\left(\frac{1}{n_1 \tilde{p}_1 (1-\tilde{p}_1)}, \frac{1}{n_2 \tilde{p}_2 (1-\tilde{p}_2)}, \dots, \frac{1}{n_m \tilde{p}_m (1-\tilde{p}_m)}\right). \quad (9)$$

A dependent variable that could be interpreted easily for purposes of directing survey crews is highly desirable and in this case, estimation of outcomes at $\mathbf{X} = \mathbf{X}_0$ is given by:

$$\hat{p}_0 = \frac{1}{1 + e^{-(\mathbf{X}_0' \hat{\mathbf{B}})}}, \quad (11)$$

\hat{p}_0 is interpreted as the probability of common crupina occurring at a given level of the regressor variable.

Separate regression models were developed for slope and aspect variables. In addition, slope and aspect models were fit at each level of vegetation class (forest, mesic grassland, or arid grassland).

For each model considered, plots of studentized residuals were examined for any departure from the standard linear regression assumptions. Models were validated using the second half of the data set. In each case, the estimated regression coefficients from the proposed model were used to predict corresponding probability values in the validation data set. Residuals from this prediction were then used for model validation.

Statistical computations were carried out using SAS/STAT (1991).

The slope, aspect, vegetation-slope, and vegetation-aspect models were used to select areas within the GIS for a hypothetical survey. Areas were selected by defining a criteria of the maximum $\hat{p}_o \pm 0.3(\hat{p}_o)$ (from eq. 11). The models were compared using as criteria, the percent of total area surveyed and the percent of total area containing common crupina that would be found using area selected by the model.

3. RESULTS AND DISCUSSION

Generalized least squares estimates for the slope and aspect models are given in Table 3. There was a quadratic relationship of common crupina occurrence to slope illustrated by the occurrence of common crupina was lower for gentle or steep slopes than it was for moderate slopes. All parameter estimates were significant at $P = 0.0001$ and were of the expected signs. A plot of studentized residuals versus slope classes showed a random and uniformly distributed arrangement of residuals (Figure 1). A plot of observed and predicted values versus slope classes indicated an adequate fit (Figure 2). Predicted values for the validation data set were calculated using the regression equation given in Table 3. Studentized residuals versus slope classes for the validation model were randomly distributed and evenly spread around zero (Figure 3).

The occurrence of common crupina was related linearly to aspect (Table 3). The occurrence of common crupina was more probable on southerly aspects than on northerly aspects. All parameters were significant at $P = 0.0001$ and had the expected signs. A plot of studentized residuals versus aspect classes showed the residuals to behave in a random fashion with no detectable pattern (Figure 4). There was a good fit of the equation to the data (Figure 5). Predicted values for the validation data set were calculated using the regression equation for aspect in Table 3. A plot of studentized residuals versus aspect class from the validation model showed an overall underestimation (Figure 6), however, most residuals were within 2 standard deviations from a mean of zero, indicating that the underestimation was not severe.

Estimated probabilities were calculated using equation 11 for both slope and aspect models. There was more than a four-fold increase in the probability of finding common crupina at a slope of 25 to 30% than there was at a slope of 0 to 2% (Figure 7). The probability of finding common crupina on steeper ground (150 to 200% slope) was half that of the probability on 25 to 30% slope. When considering aspect, common

crupina could be found three times as often on south versus north slopes (Figure 8). An attempt was made to develop a significant model combining slope and aspect but this model reduced r_i used to calculate \tilde{p}_i (eq. 3), resulting in no significant model.

Directing survey efforts using this information indicates that moderate slopes around 25 to 30% slope on southerly aspects are the best to survey. However these models do not address general changes in vegetation that occur over a region or that occur at different elevations. The slope and aspect models developed for forest, mesic grassland or arid grassland vegetation classes demonstrate the importance of a general vegetation class approach (Figure 9 & 10). In the case of slope, both forest and mesic grasslands have similar probability maximums near 24% slope but the probability maximum for arid grasslands was at 50% slope. Mesic grassland and forest classes could be treated the same when selecting the range of slopes surveyed. Arid grassland, however, would require consideration of steeper slopes than what was found for mesic grassland and forest types. The difference in the probabilities at each curve's respective maximum is a function of prevalence of common crupina within the vegetation class but also a function of the number of observations in each class. For this reason, comparison of the difference in probability between curves is not productive but the shape of the curve and location of the maximum are important for determining the range of slope classes to survey.

Aspect models for mesic grassland and arid grassland had steeper slopes than did forest (Figure 10). A narrower range of southerly aspects could be sampled in the arid and mesic grasslands than in forest.

The reason for using models that predict habitat for common crupina is to increase the efficiency of an in-field survey. The GIS was used to select areas targeted for survey based on the selection criteria established above for each model. The slope and vegetation-slope models selected 45% and 41% (Table 4) of the study area for survey. The aspect and vegetation-aspect models selected 34% (Table 4) of the study area for survey. The models containing aspect selected 7% to 11% less area for survey but 14% to 18% less common crupina would have been found. The models containing slope were comparable with slightly more common crupina found through use of the slope model (Table 4).

More arid grassland would be surveyed for both slope and aspect models than either mesic grassland or forest. Mesic grassland would be sampled less for the aspect model than for the slope model (Table 4). Considerably less common crupina (11% less) would be found in arid grassland, if a vegetation-aspect model was used instead of a vegetation-slope model.

Table 1. Slope class designations.

Slope Class	Slope Interval (%)	Slope Class	Slope Interval (%)
1	$0 \leq x < 2$	14	$35 \leq x < 40$
2	$2 \leq x < 4$	15	$40 \leq x < 45$
3	$4 \leq x < 6$	16	$45 \leq x < 50$
4	$6 \leq x < 8$	17	$50 \leq x < 55$
5	$8 \leq x < 10$	18	$55 \leq x < 60$
6	$10 \leq x < 12$	19	$60 \leq x < 70$
7	$12 \leq x < 14$	20	$70 \leq x < 80$
8	$14 \leq x < 16$	21	$80 \leq x < 90$
9	$16 \leq x < 18$	22	$90 \leq x < 100$
10	$18 \leq x < 20$	23	$100 \leq x < 150$
11	$20 \leq x < 25$	24	$150 \leq x < 200$
12	$25 \leq x < 30$	25	≥ 200
13	$30 \leq x < 35$		

Table 2. Aspect class designations.

Aspect Class	Aspect Interval (degree)
1	0 to 9, 351 to 360
2	10 to 19, 341 to 350
3	20 to 29, 331 to 340
4	30 to 39, 321 to 330
5	40 to 49, 311 to 320
6	50 to 59, 301 to 310
7	60 to 69, 291 to 300
8	70 to 79, 281 to 290
9	80 to 89, 271 to 280
10	90 to 99, 261 to 270
11	100 to 109, 251 to 260
12	110 to 119, 241 to 250
13	120 to 129, 231 to 240
14	130 to 139, 221 to 230
15	140 to 149, 211 to 220
16	150 to 159, 201 to 210
17	160 to 169, 191 to 200
18	170 to 180, 181 to 190

Table 3. Estimated regression coefficients for slope and aspect models.

Term	Parameter Estimate	Standard Error	t	P> t	R ²
Slope					
Intercept	-2.8498	0.1234	-23.1	0.001	0.89
Linear	0.2872	0.0232	12.4	0.001	
Quadratic	-0.0108	0.0001	-10.9	0.001	
Aspect					
Intercept	-2.2837	0.0904	-25.3	0.001	0.87
Linear	0.0807	0.0078	10.4	0.001	

Table 4. Contrasting slope and aspect model survey efficiency. Column 3 is the percent of the area within a vegetation class that was selected for survey. Column 4 is the percent of the study area selected for survey. The percent of common crupina within a vegetation class that would have been found during a survey is in column 5. In column 6 the percent of common crupina found is expressed as a percent of the total amount of common crupina in the study area.

Model	Vegetation Class	Area Surveyed		Common Crupina Found	
		(% within class)	(% of total)	(% within class)	(% of total)
Slope	Forest	23	7	63	20
	Mesic Grass.	65	12	75	8
	Arid Grass.	39	<u>22</u>	74	<u>42</u>
	(total =)		41		70
Aspect	Forest	40	9	33	19
	Mesic Grass.	30	6	54	6
	Arid Grass.	34	<u>19</u>	55	<u>31</u>
	(total =)		34		56
Slope	N/A	N/A	45	N/A	74
Aspect	N/A	N/A	34	N/A	56

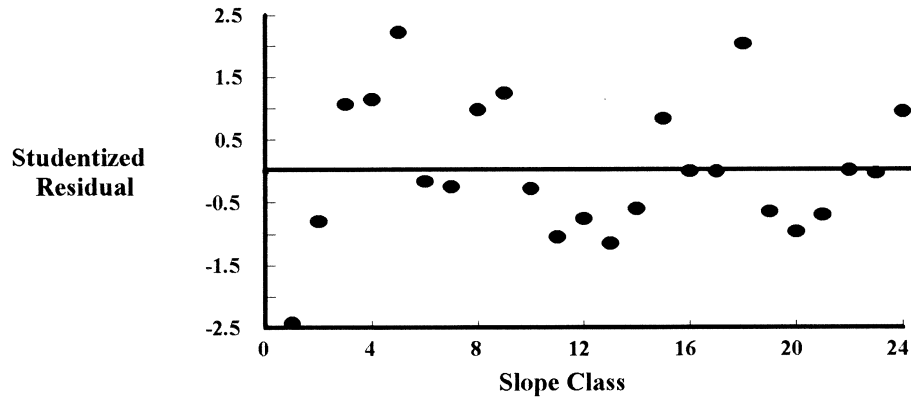


Figure 1. Plot of studentized residuals versus slope classes.

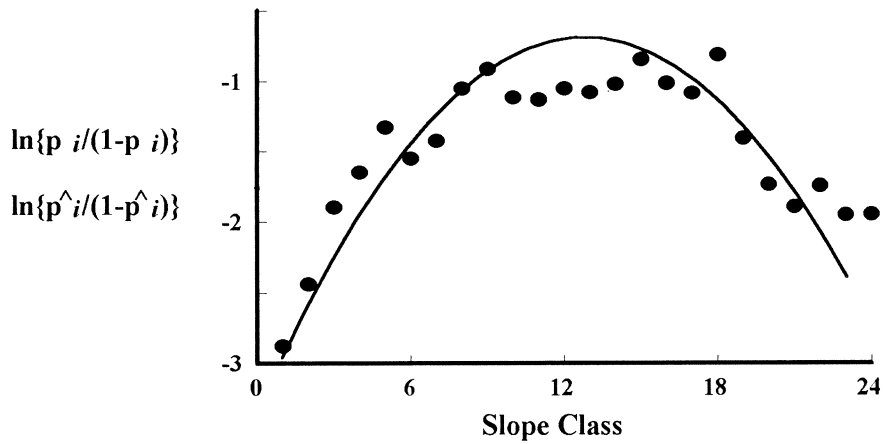


Figure 2. Plot of predicted and observed values versus slope classes.

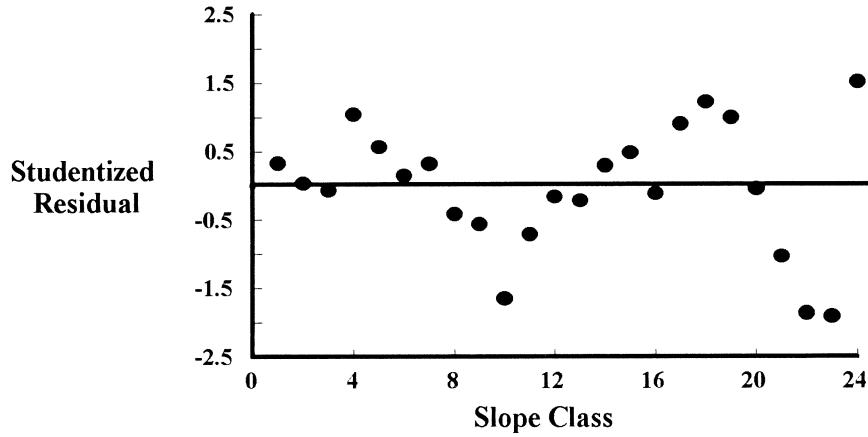


Figure 3. Plot of studentized residuals versus slope classes for the validation model.

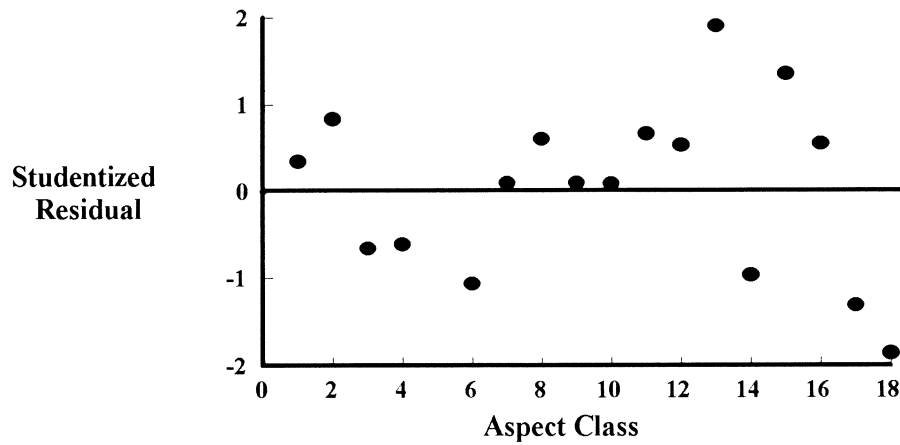


Figure 4. Plot of studentized residuals versus aspect classes.

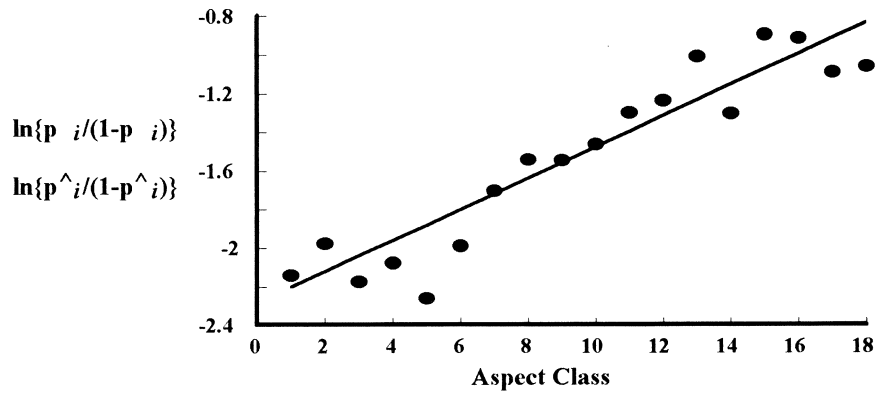


Figure 5. Plot of predicted and observed values versus aspect classes.

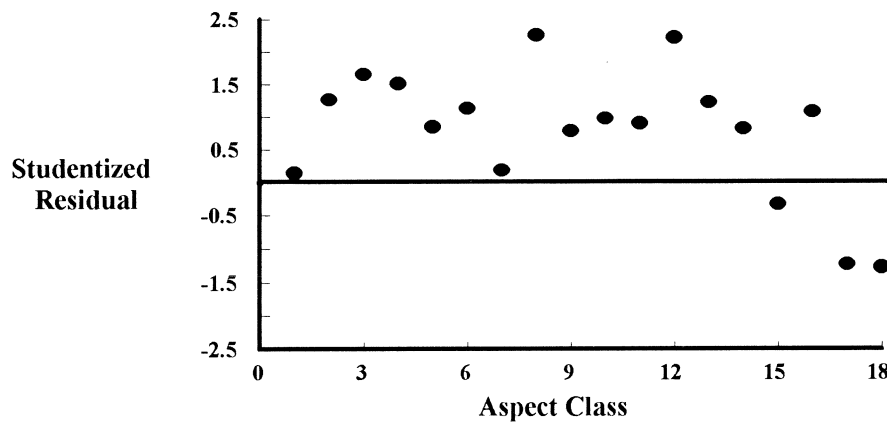


Figure 6. Plot of studentized residuals versus aspect classes for the validation model.

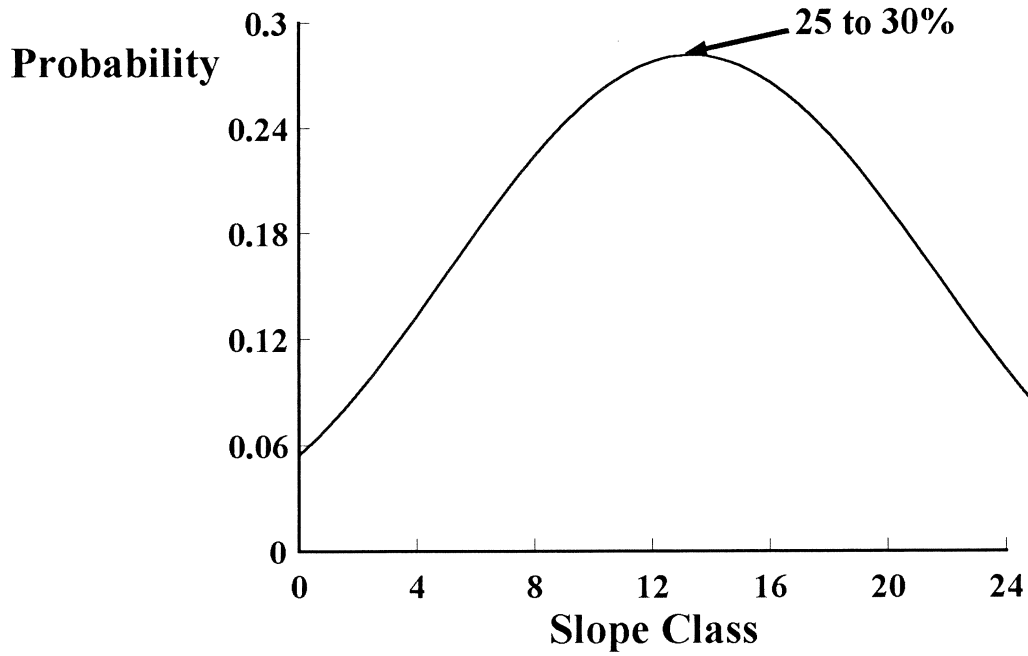


Figure 7. Estimated probabilities associated with slope classes.

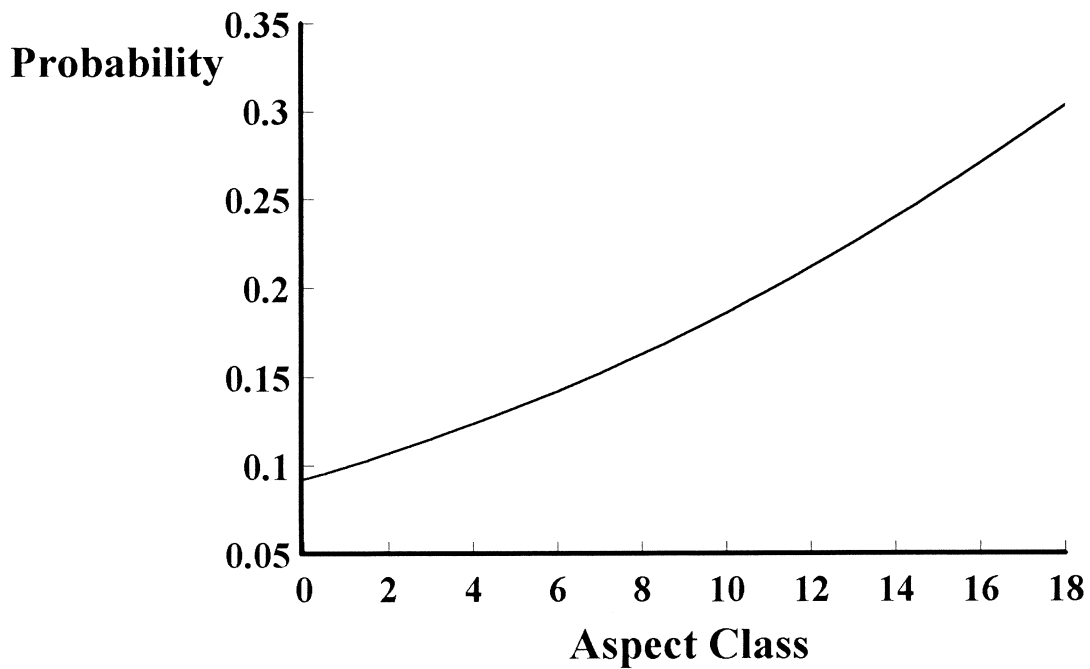


Figure 8. Estimated probabilities associated with aspect classes.

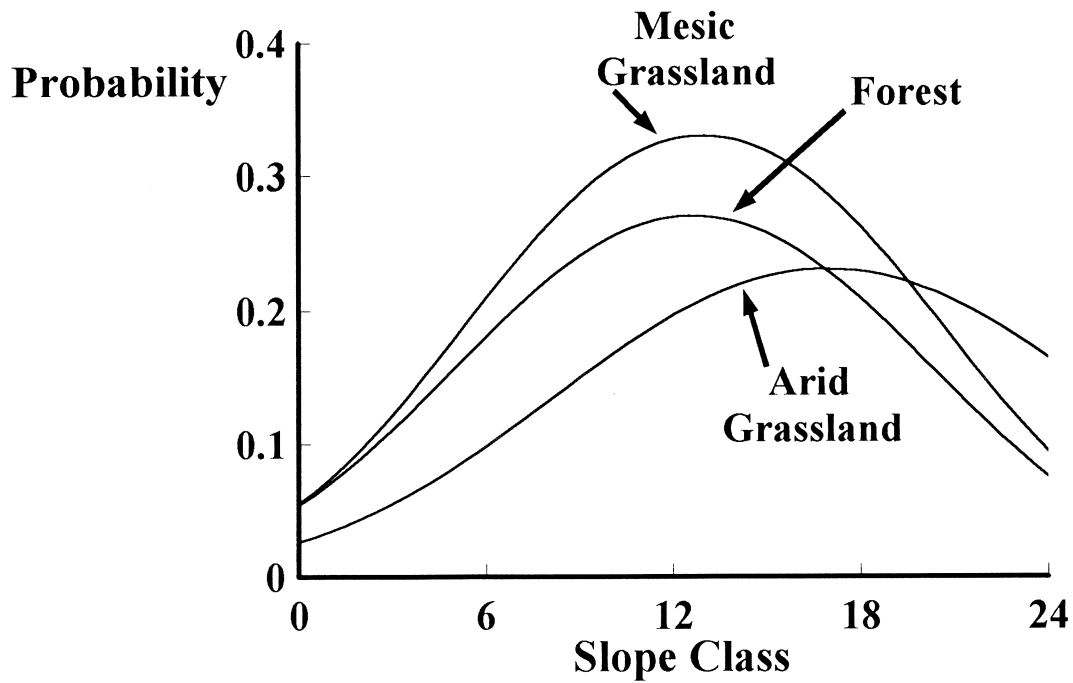


Figure 9. Estimated probabilities associated with vegetation and slope classes.

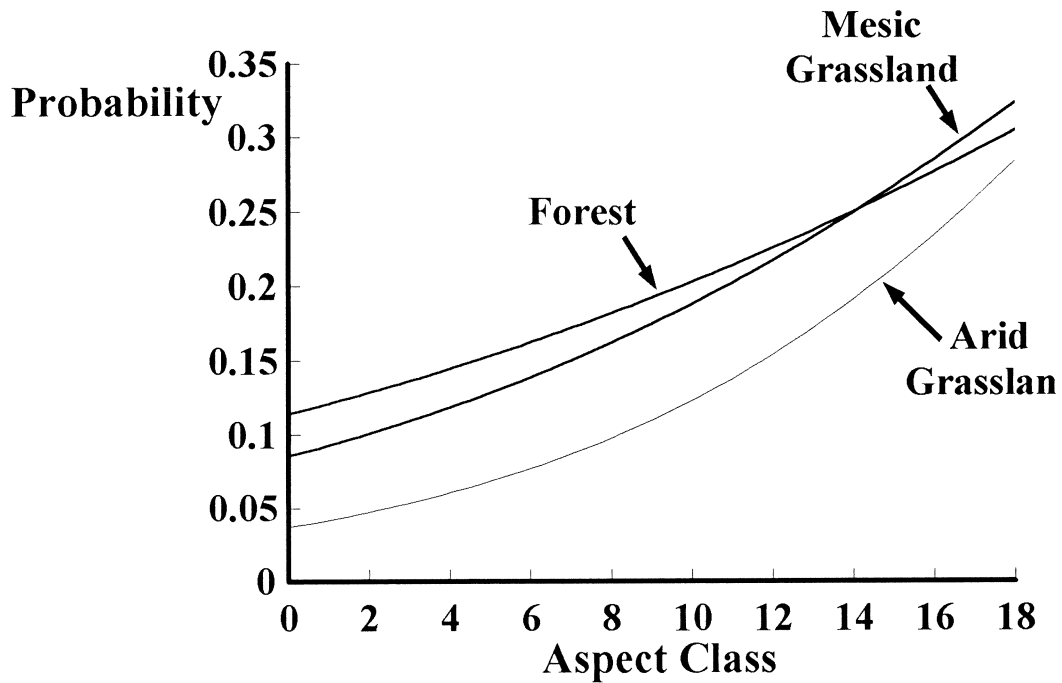


Figure 10. Estimated probabilities associated with vegetation and aspect classes.

4. SUMMARY

Logistic regression was used successfully to predict common crupina occurrence. All model parameters were highly significant and validation of the models using a second data set was accomplished. Using these models to direct survey efforts would limit the area surveyed to 34 to 42% of the total area. Fitting models of slope or aspect for each vegetation class decreased area surveyed for slope models (3% less area) but there was a subsequent 4% decrease in the percent of common crupina found. There were no differences in area surveyed or common crupina found when comparing the aspect and vegetation-aspect models. Overall, the general aspect and slope models performed as well as the vegetation class specific slope and aspect models. In regions containing large areas of forested land, the vegetation-slope model would decrease area surveyed and find a higher percentage of common crupina.

REFERENCES

1. Prather, T. S. and R. H. Callihan. 1993. Weed eradication using geographic information systems. *Weed Technol.* 7:265-269.
2. SAS Institute. 1991. SAS/STAT users guide. SAS Institute, Inc., Cary, NC.
3. Zamora, D. L., D. C. Thill, and R. E. Eplee. 1989a. An eradication plan for plant invasions. *Weed Technol.* 3:2-12.
4. Zamora, D. L., D. C. Thill, and R. E. Eplee. 1989. Eradication manual for common crupina (*Crupina vulgaris*). Univ. Id., Coll. Agric., Bull. No. 701.