

Kansas State University Libraries

**New Prairie Press**

---

Conference on Applied Statistics in Agriculture

1999 - 11th Annual Conference Proceedings

---

## ESTIMATING THE LIKELIHOOD OF YELLOW STARHISTLE OCCURRENCE USING AN EMPIRICALLY DERIVED NONLINEAR REGRESSION MODEL

Bahman Shafii

William J. Price

Lawrence W. Lass

Donn C. Thill

*See next page for additional authors*

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

Shafii, Bahman; Price, William J.; Lass, Lawrence W.; and Thill, Donn C. (1999). "ESTIMATING THE LIKELIHOOD OF YELLOW STARHISTLE OCCURRENCE USING AN EMPIRICALLY DERIVED NONLINEAR REGRESSION MODEL," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1257>

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](mailto:cads@k-state.edu).

---

**Author Information**

Bahman Shafii, William J. Price, Lawrence W. Lass, and Donn C. Thill

## ESTIMATING THE LIKELIHOOD OF YELLOW STARHISTLE OCCURRENCE USING AN EMPIRICALLY DERIVED NONLINEAR REGRESSION MODEL

Bahman Shafii and William J. Price  
Statistical Programs  
Lawrence W. Lass and Donn C. Thill  
Division of Plant Sciences

College of Agriculture  
University of Idaho  
Moscow, Idaho 83844

### ABSTRACT

Yellow starthistle is a noxious weed common in the semiarid climate of Central Idaho and other western states. Early detection of yellow starthistle and predicting its infestation potential have important scientific and managerial implications. Weed detection and delineation are often carried out by visual observation or survey techniques. However, such methods may be ineffective in detecting sparse infestations. The distribution of yellow starthistle over a large region may be affected by various exogenous variables such as elevation, slope and aspect. These landscape variables can be used to develop prediction models to estimate the potential invasion of yellow starthistle into new areas. A nonlinear prediction model has been developed based on a polar coordinate transformation to investigate the ability of landscape characteristics to predict the likelihood of yellow starthistle occurrence in North Central Idaho. The study region included the lower Snake river and parts of the Salmon and Clearwater basins encompassing various land use categories. The model provided accurate estimates of incidence of yellow starthistle within each specified land use category and performed well in subsequent statistical validations.

### INTRODUCTION

Yellow starthistle (*Centaurea solstitialis* L.) is an introduced noxious weed currently infesting millions of acres of rangeland in the United States. It is considered poor forage for all classes of livestock and may cause a fatal neurological disorder in horses (Sheley et al. 1999). It can also cause serious economic loss due to its potential forage yield reduction and ecosystem degradation of grasslands.

Yellow starthistle is common in the semiarid climate of northern Idaho and many other western states. It thrives best on warm, deep, well drained soils with 30-75 cm of annual

precipitation. However, this winter annual can survive and dominate annual plant communities in unproductive soils (e.g. rocky and shallow sites) when annual precipitation is below 25 cm. Major invasion has occurred on rangeland and non-crop land, however, cultivated lands such as dryland grain, grass, legume, seed crop and pasture are also susceptible to invasion by yellow starthistle (Lass et al., 1999).

Yellow starthistle is native to the east central region of Europe, and its seeds are believed to have been brought in alfalfa seeds shipped to North America in the early 1800's (Sheley et al., 1999). It is estimated that nearly two million ha of land are currently infested with yellow starthistle in the Western States (Lass et al., 1999). Infestations have been established in Idaho, Oregon, California, Washington, and Utah, and most recently in Arizona, Colorado, Montana, Nevada, New Mexico, and North Dakota. In 1955, the infestation in Idaho was less than 10 ha, but has increased to approximately 200,000 ha in less than 45 years.

Early detection of yellow starthistle and predicting its infestation potential is an important consideration as the plant expands into new areas. While identification of weed infestations is carried out using visual observations, such methods are often ineffective in detecting sparse infestations in remote areas. Prediction models will allow land managers to focus on sites with a high likelihood of infestation based on demographic characteristics (e.g. elevation, slope, land use) of currently infested sites. Landscape variables such as slope, aspect, and elevation have been used previously to determine the likelihood of occurrence for a specified weed (e.g. Prather and Shafii, 1994; Dewey et al., 1991) or vegetation species (e.g. Myster et al., 1997).

The objective of this research was to develop an empirically derived prediction model based on landscape characteristics to assess the likelihood of yellow starthistle occurrence in North Central Idaho.

## METHODS

Several candidate models were considered to investigate the potential relationship between incidence of yellow starthistle and the landscape variables slope and aspect. In all model formulations, let

$$\begin{aligned} A_i &= \text{the } i^{\text{th}} \text{ aspect,} \\ S_j &= \text{the } j^{\text{th}} \text{ slope, and} \\ Z_{ij} &= \text{logit} = \ln(p_{ij}/(1-p_{ij})), \end{aligned}$$

where  $p_{ij}$  is the proportion of yellow starthistle present at the  $i$ th level of aspect and  $j$ th level of slope,  $i = 1, 2, 3, \dots, k$  and  $j = 1, 2, 3, \dots, n$ . These models may be grouped into the following four categories:

Model I: Two-segment quadratic in aspect

$$\begin{aligned} Z_{ij} &= \beta_{01} + \beta_{11}A_i + \beta_{21}A_i^2 + \beta_{31}S_j + \varepsilon_{ij} ; \text{ for } A_i < m \\ Z_{ij} &= \beta_{02} + \beta_{12}A_i + \beta_{22}A_i^2 + \beta_{32}S_j + \varepsilon_{ij} ; \text{ for } A_i \geq m \end{aligned}$$

where  $m$  is the value of aspect where the segments meet (joint point).

Model II: Two-segment quadratic in aspect and slope

$$Z_{ij} = \beta_{01} + \beta_{11}A_i + \beta_{21}A_i^2 + \beta_{31}S_j + \beta_{41}S_j^2 + \varepsilon_{ij} ; \text{ for } A_i < m$$

$$Z_{ij} = \beta_{02} + \beta_{12}A_i + \beta_{22}A_i^2 + \beta_{32}S_j + \beta_{42}S_j^2 + \varepsilon_{ij} ; \text{ for } A_i \geq m$$

where, as in Model I,  $m$  is the joint point on the aspect axes.

Model III: Six-segment quadratic in aspect

$$Z_{ij} = \beta_{01} + \beta_{11}A_i + \beta_{21}A_i^2 + \varepsilon_{ij} ; \text{ for } A_i < m ; S_j < l_1$$

$$Z_{ij} = \beta_{02} + \beta_{12}A_i + \beta_{22}A_i^2 + \varepsilon_{ij} ; \text{ for } A_i < m ; l_1 \leq S_j < l_2$$

$$Z_{ij} = \beta_{03} + \beta_{13}A_i + \beta_{23}A_i^2 + \varepsilon_{ij} ; \text{ for } A_i < m ; S_j > l_2$$

$$Z_{ij} = \beta_{04} + \beta_{14}A_i + \beta_{24}A_i^2 + \varepsilon_{ij} ; \text{ for } A_i \geq m ; S_j < l_1$$

$$Z_{ij} = \beta_{05} + \beta_{15}A_i + \beta_{25}A_i^2 + \varepsilon_{ij} ; \text{ for } A_i \geq m ; l_1 \leq S_j < l_2$$

$$Z_{ij} = \beta_{06} + \beta_{16}A_i + \beta_{26}A_i^2 + \varepsilon_{ij} ; \text{ for } A_i \geq m ; S_j > l_2$$

where  $m$  is the joint point on the aspect axes, and  $l_1$  and  $l_2$  are joint points on the slope axes.

Model IV: Polar coordinate model

$$Z_{ij} = \beta_0 + \beta_1X + \beta_2X^2 + \beta_3Y + \beta_4Y^2 + \beta_5\ln(X^2 + Y^2) + \varepsilon_{ij}$$

where  $X$  and  $Y$  are polar coordinate transformations, i.e.:  $X = \text{slope} \cdot \cos(\text{aspect})$  and  $Y = \text{slope} \cdot \sin(\text{aspect})$ .

For all models,  $\beta$ 's are unknown regression coefficients and for each specified model,  $\varepsilon_{ij}$  is assumed  $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$ . In models I-III, the joint points are determined via minimum mean squares error criterion. The reasons behind various model formulations as well as the rationale for considering polar transformations will be discussed in the Empirical Results section.

Estimation of parameters in the specified models was accomplished using a weighted nonlinear regression algorithm. Statistical validation was carried out using the method of data splitting (Snee, 1977). Here, the data are partitioned into two subsamples, a fitting sample and a validation sample. The fitted model is subsequently used to estimate the response in the validation sample. The residual structure for the validation sample is then examined for unusual patterns or trends in order to assess the model adequacy.

Statistical computations and graphics were performed using SAS/STAT (1991a) and SAS/GRAPH (1991b), respectively.

## EMPIRICAL RESULTS

### Source and Description of Data

The study region covered the lower Snake river and parts of the Salmon and Clearwater river basins of North Central Idaho. The boundary of the study area was latitude  $45^{\circ}$  N, Longitude  $-118^{\circ}$  E and Latitude  $48^{\circ}$  N, Longitude  $-114^{\circ}$  E. This area encompasses various land use categories over a wide range of topography including flat river bottoms, steep canyons, and high mesas. Specific data sets used were regional USGS 1 degree digital elevation models (DEM) for elevation, slope and aspect, USGS Landsat-based land use classification data, and a presence/absence survey of yellow starthistle conducted by the University of Idaho. Area analysis of the land use data showed 24 % crop and pasture land, 22% mixed rangeland, and 50% evergreen forest. The remaining 4% consisted of a wide array of land uses such as water, urban, barren land, etc., and was considered negligible for inclusion in the study.

The DEM, land use and yellow starthistle survey data were aligned to a common grid size of 80 x 80 m resulting in over 3.3 million grid cells. Prior to analysis, this data set was randomly split into two equally sized portions for the purposes of modeling and validation. Each portion was then subdivided according to the three aforementioned land use categories. All the procedures described below were carried out separately for each specified land use category, however, due to limited space, only the details of the crop and pasture category will be presented here.

The landscape variables considered as potential regressors were elevation, slope and aspect. Initial investigation showed elevation to be of little value in predicting the likelihood of yellow starthistle infestation, and therefore efforts were concentrated on slope and aspect variables. In order to obtain multiple observations in a tabulation of slope and aspect, these variables were categorized as follows: slope, ranging from 0 to 50%, was divided every 5% resulting in 10 categories with class 1 = flat and class 10 = steep; aspect, ranging from 0 to 359 degrees, was divided every 22.5 degrees in a clock-wise direction resulting in 16 categories with class 1 = north. The choice for these specific categories was made based on previous experience and limiting resolutions of the data. The resulting cross classification was a 10 by 16 table with each cell containing the proportion of yellow starthistle presence. A logit transformation was carried out on the proportions in order to mitigate variability and meet distributional assumptions.

Figure 1 presents the observed logits for the crop and pasture land use. One of the more prominent features of the data is seen in Figure 1a. The response across aspect is cyclic in nature showing low values in the northeast direction (aspect classes 3-5), high values in the southwest (aspect classes 9-11) and low values in the northwest direction (aspect classes 13-15). Thus, the highest yellow starthistle presence seems to be in the 180 to 220 degree aspect. Figure 1b depicts observed logits versus slope classes. Here a convex shape is seen, peaking around 25 to 30% (slope classes 5-6). There is also a noticeable increase in variability as slope increases. As slopes become steeper, the number of available data values decreases, leading to less precision at these points. For this reason, a weighted least squares procedure, with weights proportional to sample sizes, was used for all model estimations. Figure 1c gives the full three-dimensional relationship of yellow starthistle with slope and aspect. The cyclic nature across aspect can now be seen to vary with slope, becoming more pronounced at steeper inclines. The jagged surface at high slopes

is due to sparse data and reduced precision.

### Modeling the Likelihood of Yellow Starthistle Occurrence

As a first attempt to assess the incidence of yellow starthistle, Model I was used. The estimated joint point along with parameter estimates are given in Table 1a. The initial goal was to capture the cyclic characteristic of the aspect axis with a two-segment quadratic surface (Figure 2). While this proved successful at lower slope values, serious deviation from the data occurred at higher slope values where the linearity of the slope term failed to account for the bending along that axis. The residuals in Figure 2b clearly indicate this pattern.

Parameter estimates for Model II are provided in Table 1b. They are significant ( $p < 0.05$ ) and the joint point is estimated to be  $m=9$ . The addition of a quadratic slope term in Model II, was an attempt to address the bending along that axis. This resulted in the correct modeling of the convex shape along the slope axis, however, the prominent oscillation at high slope values was not evident in the fitted model (Figure 3a). Residual plots also indicate very large studentized values in some areas (Figure 3b).

In order to allow maximum flexibility of the fitted surface, Model III was designed with six segments, each of which was a quadratic form in aspect. The estimated joint point for the aspect axis was  $m=9$  and those for slope were  $l_1=2$  and  $l_2=7$ . Most of the 18 parameters were significant (Table 1c). The model did a good job predicting the wavy surface of observed data (Figure 4a). The convex shape across slope was adequately modeled using three segments while the freedom of multiple quadratic terms allowed a better representation of the oscillations across aspect. Residuals were of moderate size and showed minimal patterns (Figure 4b). To achieve these results, however, required 18 parameters in a large segmented model.

Reassessment of the problem showed that the specified cyclic pattern might have been related to the values assigned to the aspect classes (Figure 1a). In this case a value of 1 was arbitrarily assigned to zero degrees. Altering this to any of the other 358 possible configurations would significantly modify the observed surface given in Figure 1c. Consequentially, this would result in different parameter estimates and possibly even necessitate respecification of the model.

To better account for the cyclic nature of aspect, a change of coordinate systems was considered. This was accomplished by mapping the data to a polar coordinate system where the angle of aspect ( $\theta$ ) and the value of slope ( $r$ ) were used to define two new general axes,  $x=r*\cos(\theta)$  and  $y=r*\sin(\theta)$  (Figure 5). Such transformations have been used previously to develop models based on landscape variables (Stage, 1976).

The transformed data is shown in Figure 6. The logit response shows a broad curving or fanning pattern across the x axis with a noticeable dip towards  $x=0$ . On the y axis, the response is slightly concave with a negative slant. It too has an obvious dip at the origin,  $y=0$ .

The basis for Model IV starts with a parabolic, possibly saddle shaped, quadratic surface to account for the curvature in x and y. The addition of linear terms provides for the slanting of the surface. Finally, a logarithmic component proportional to the radius, i.e. slope, is added to pull a funnel shape down through the quadratic surface toward the origin.

The parameter estimates and standard errors of Model IV are given in Table 1d. All parameter estimates were significant ( $p < 0.05$ ). The quadratic surface in X, in combination with the logarithmic term in radius, provide sufficient curvature to account for the concavity observed

on the Y axis. This can be seen from the close agreement between predicted and observed values plotted on the polar axes (Figure 7). The fanning of the x axis, the slanting concavity of the y axis and the dip towards the axes center have all been accurately modeled. The corresponding plots for the untransformed data are given in Figure 8. The logit oscillation across aspect appears to be correctly predicted in the model (Figure 8a). Minima in the northeast and northwest directions are present as is the maximum in the southwest direction. The curvature along the slope axis is also predicted with logit values spreading out as slope increases (Figure 8b).

A comparative view of the predicted surface (Figure 9a) shows close correspondence to the observed values. As slope increases the cyclic response to aspect becomes more pronounced. The overall convex shape of slope is also present. Residual plots in Figure 9b have minimal magnitude and conform to expected patterns.

The transformation of axes resulted in a parsimonious model which predicted characteristics of the observed data well and had the smallest Mean Square Error of all the models tested (Model IV: 6 parameters,  $MSE=0.02$ ). In contrast, segmented models provided minimal fits (Model I : 8 parameters,  $MSE=114.8$ ; Model II, 10 parameters,  $MSE=52.19$ ) or unacceptable solutions (Model III: 18 parameters,  $MSE=0.04$ ).

Given the above results, further testing of Model IV was carried out using cross validation. Residual values obtained by differencing the observed validation data from the predicted model gave very reasonable results. Residuals along polar axes showed only slight patterns and no excessive magnitudes (Figure 10a). When viewed on the untransformed axes, the residuals were comparable to the fitted data with no detectable trends and minimal values.

Modeling efforts on the other land use categories using the polar coordinate transformation also proved successful. The observed data exhibited similar patterns, and the use of the polar coordinate approach with Model IV provided very close approximations. Cross validation of the estimated models for each of the land use categories produced acceptable results. Taken as a group, the specified models provided an accurate assessment of the yellow starthistle infestation potential for the study area.

## CONCLUDING REMARKS

Predicting the likelihood of yellow starthistle incidence and infestation potential have important managerial implications. Prediction will allow land managers to concentrate their efforts on sites with high likelihood of yellow starthistle occurrence based upon specified landscape characteristics. This is especially true with current survey techniques that are often ineffective in identifying sparse infestations. A model was developed based on landscape characteristics, slope and aspect, which demonstrated exceptional ability in predicting the likelihood of yellow starthistle incidence in North Central Idaho. The model was more parsimonious than its polynomial counterparts, provided more accurate prediction within each of the land use categories, and tested well in subsequent statistical validations.



## REFERENCES

- Dewey, S. A., K. P Price and D. Ramsey. 1991. Satellite remote sensing to predict potential distribution of Dyers Woad. *Weed Technology* 5:479-484.
- Lass, L. W. J. P. McCaffrey, D. C. Thill and R. H. Callihan. 1999. Yellow starthistle biology and management in pasture and rangeland. *University of Idaho Bulletin No. 805*, 18 pp.
- Myster, R. W., J. R. Thomlinson, and M. C. Larsen. 1997. Predicting landslide vegetation in patches on landscape gradients in Puerto Rico. *Landscape Ecology*, 12: 299-307.
- Prather, T. S. and B. Shafii. 1994. Predicting common crupina habitat with geographic and remote sensing data. *Proceedings of the 1994 Kansas State University Conference on Applied Statistics in Agriculture*. Kansas State University, Manhattan, Ks, 122-135.
- SAS Institute Inc. 1991a. *SAS/STAT User's Guide, Version 6 Fourth Ed., Vol. 2*. Sas Institute Inc, Cary NC.
- SAS Institute Inc. 1991b. *SAS/GRAPH User's Guide, Version 6 Fourth Ed., Vol. 2*. Sas Institute Inc, Cary NC.
- Sheley, R. L., L. L. Larson, and J. S. Jacobs. 1999. Yellow Starthistle *In* Roger L. Sheley and Janet K. Petroff (eds.) : *Biology and Management of Noxious Rangeland Weeds*. Oregon State University Press, Corvallis, OR., pp 408-416.
- Snee, R. D. 1977. Validation of regression models: Methods and Examples. *Technometrics*, 19: 415-428.
- Stage, A. R. 1976. An expression for the effect of aspect, slope and habitat type on tree growth. *Forest Science*, Vol. 22, No. 4, pp 457-460.

**Table 1.** Parameter estimates, standard errors, and estimated joint points for Model I (a), Model II (b), Model III (c) and Model IV (d).

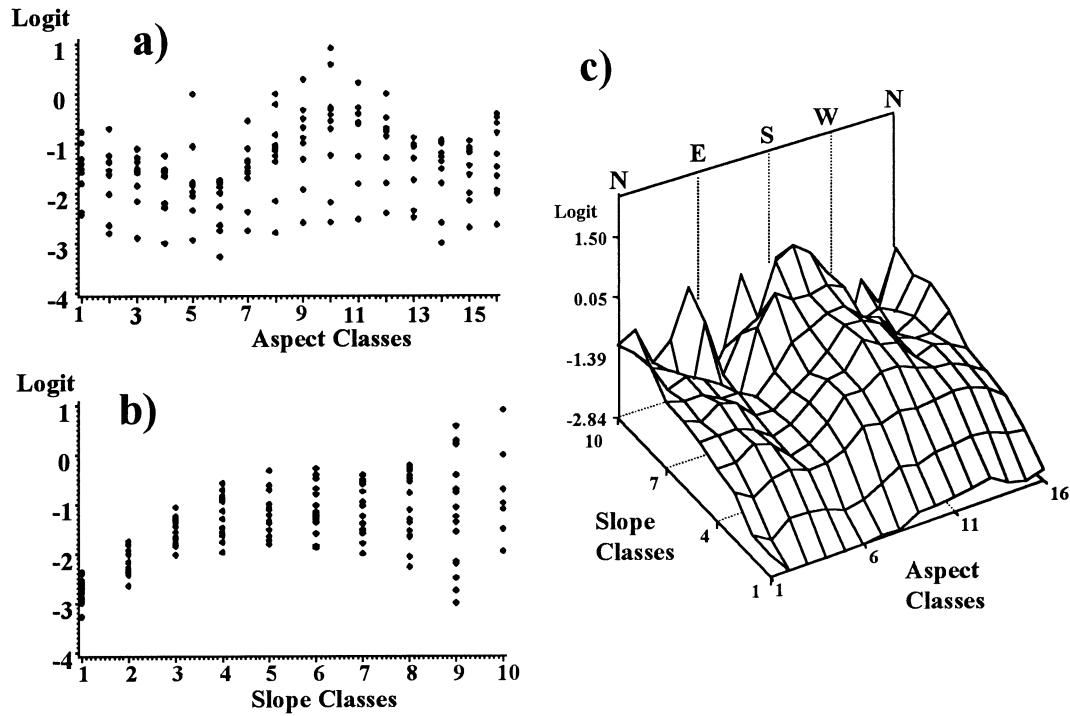
a) Model I			b) Model II		
Parameter	Estimate	Std Error	Parameter	Estimate	Std Error
$\beta_{01}$	-2.55	0.102	$\beta_{01}$	-2.91	0.080
$\beta_{11}$	-0.26	0.048	$\beta_{11}$	-0.29	0.033
$\beta_{21}$	0.03	0.005	$\beta_{21}$	0.03	0.003
$\beta_{31}$	0.31	0.019	$\beta_{31}$	0.70	0.045
$\beta_{02}$	-5.15	1.880	$\beta_{41}$	-0.06	0.006
$\beta_{12}$	0.42	0.293	$\beta_{02}$	-5.99	1.270
$\beta_{22}$	-0.02	0.011	$\beta_{12}$	-0.47	0.198
$\beta_{32}$	0.37	0.023	$\beta_{22}$	-0.02	0.008
			$\beta_{32}$	0.89	0.055
			$\beta_{42}$	-0.07	0.008

m = 9

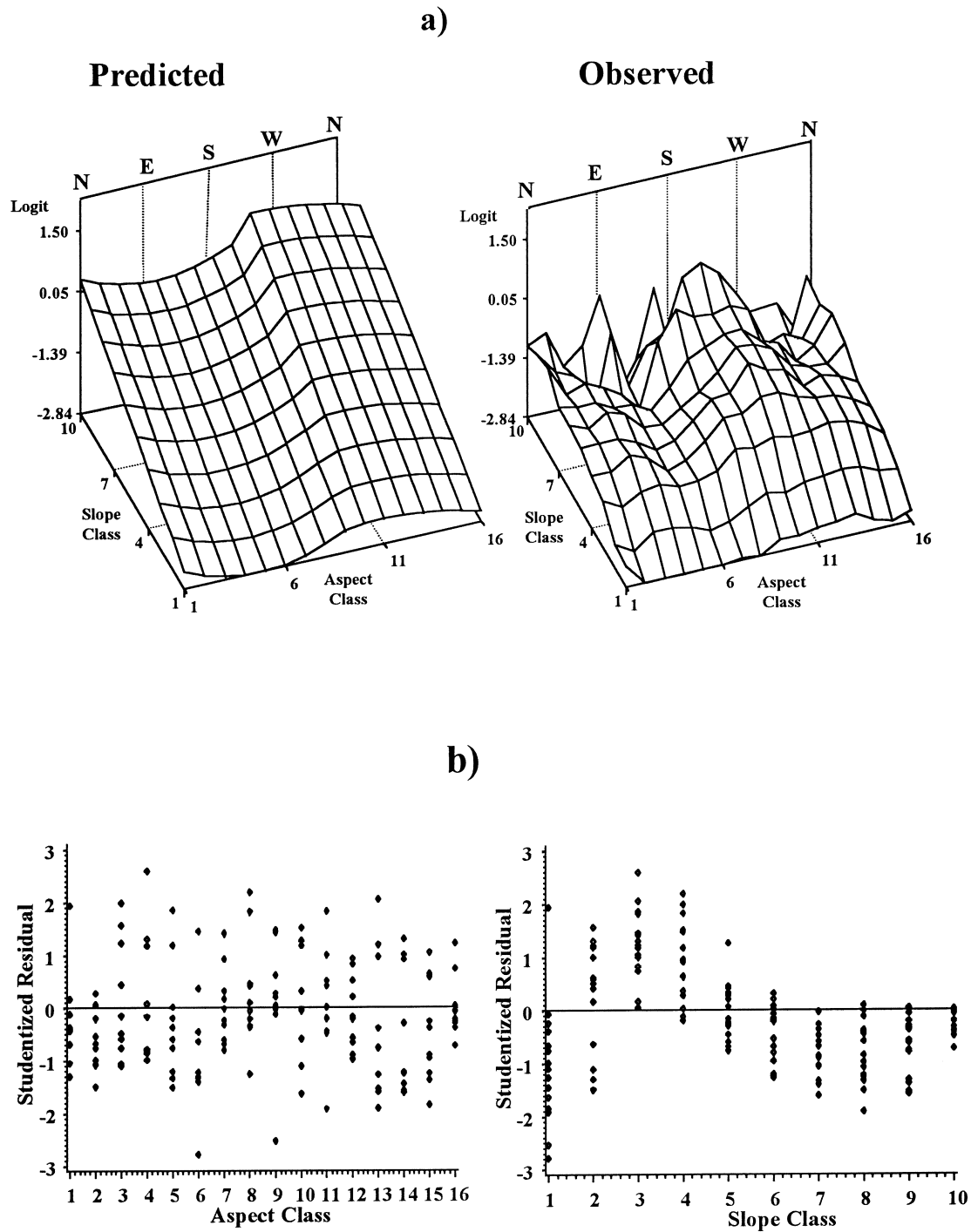
  

c) Model III			d) Model IV		
Parameter	Estimate	Std Error	Parameter	Estimate	Std Error
$\beta_{01}$	-2.18	0.201	$\beta_0$	-0.21	0.103
$\beta_{11}$	-0.27	0.092	$\beta_1$	0.24	0.079
$\beta_{21}$	0.03	0.009	$\beta_2$	1.47	0.211
$\beta_{02}$	-1.24	0.235	$\beta_3$	-0.77	0.077
$\beta_{12}$	-0.24	0.108	$\beta_4$	0.40	0.199
$\beta_{22}$	0.03	0.011	$\beta_5$	0.55	0.026
$\beta_{03}$	-0.13	0.434			
$\beta_{13}$	-0.71	0.207			
$\beta_{23}$	0.08	0.021			

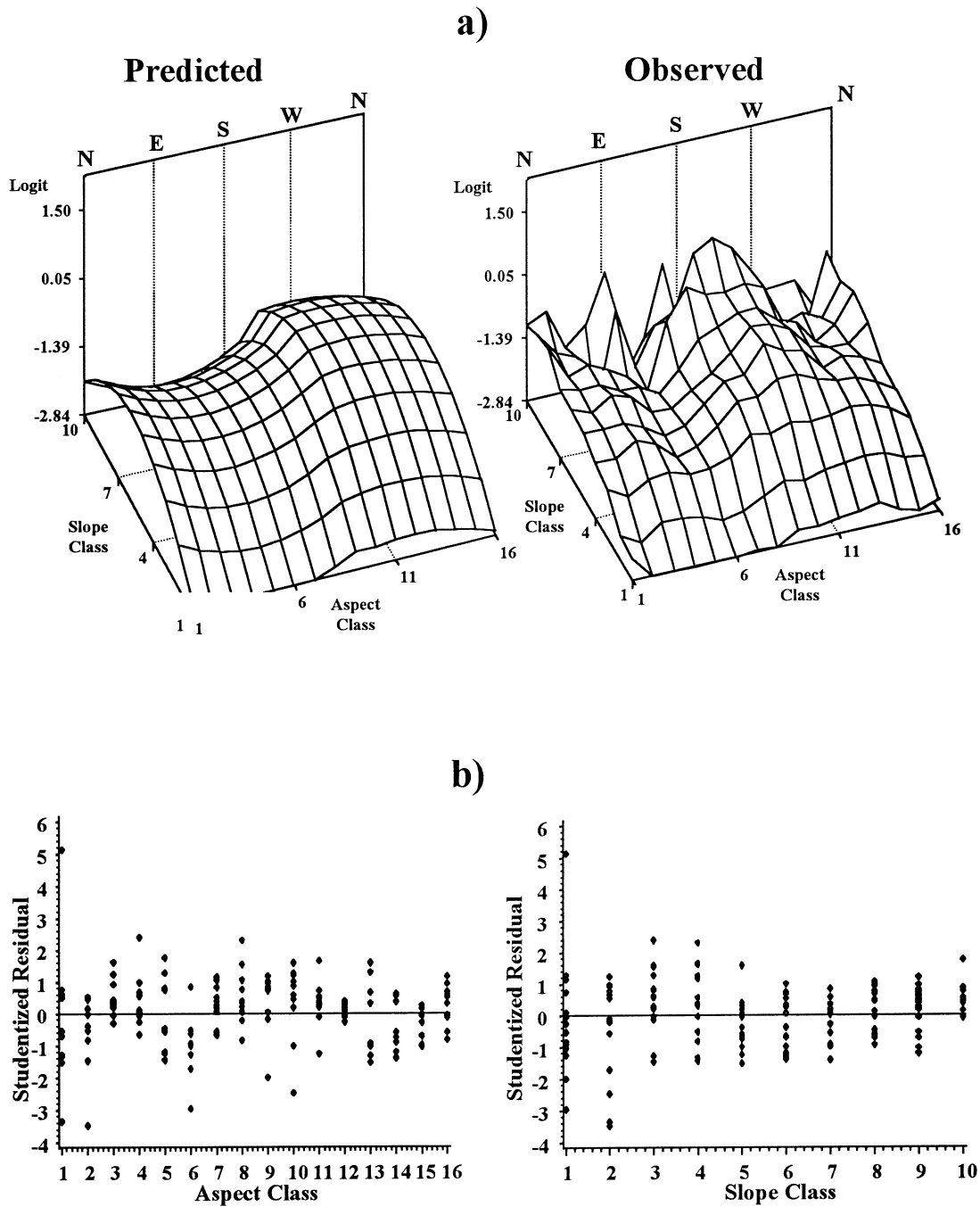
m=9



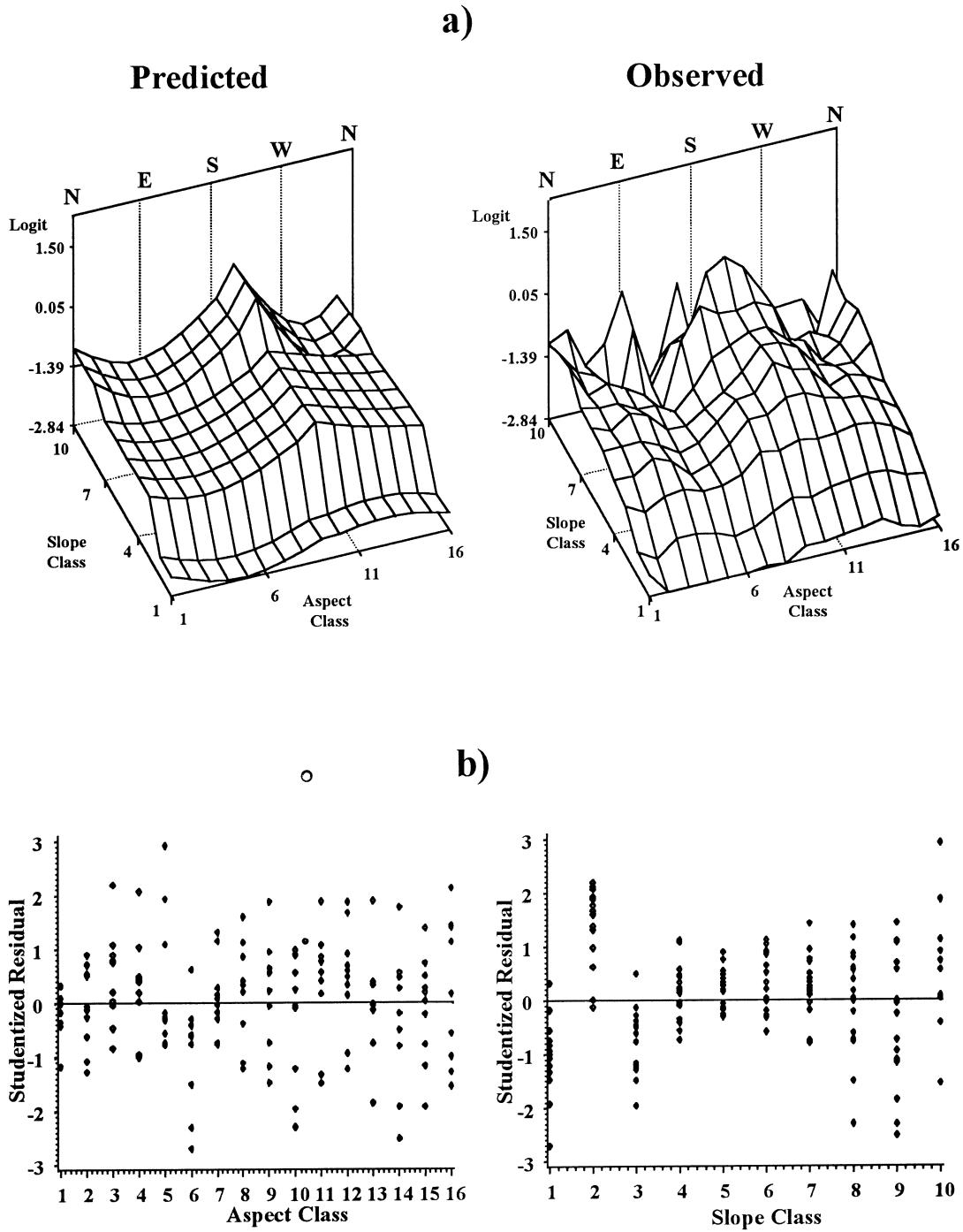
**Figure 1.** Observed Logits for Aspect (a) and Slope (b) classes along with the corresponding three dimensional surface (c).



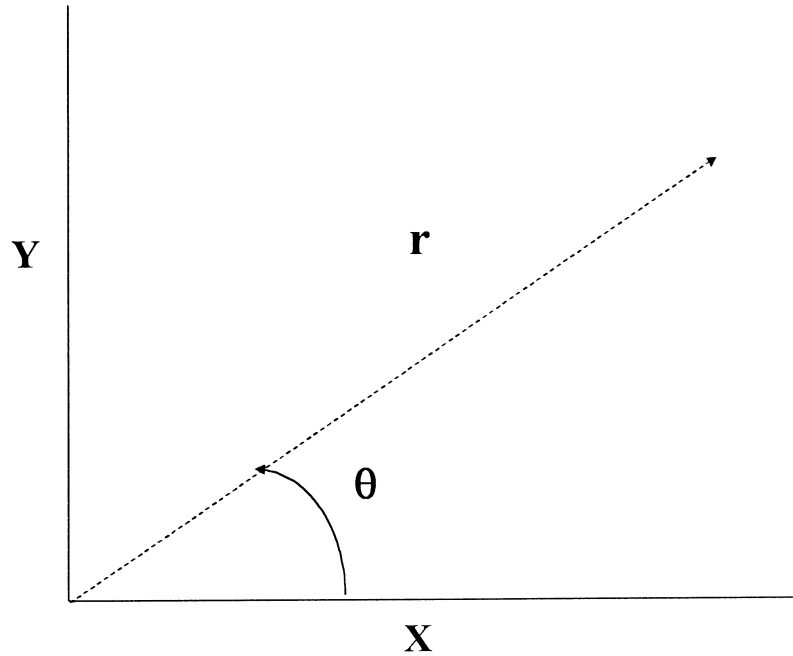
**Figure 2.** Predicted and observed surfaces for Model I (a) along with the underlying residual structure (b).



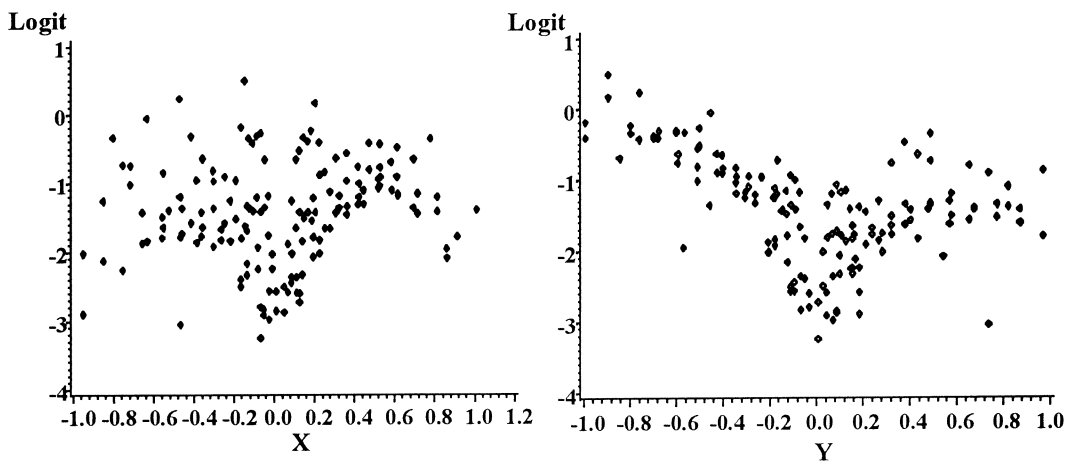
**Figure 3.** Predicted and observed surfaces for Model II (a) along with the underlying residual structure (b).



**Figure 4.** Predicted and observed surfaces for Model III (a) along with the underlying residual structure (b).



**Figure 5.** Polar coordinate conversion with  $\theta$  and  $r$  representing aspect (radians) and radius (%), respectively, and  $X=r*\cos(\theta)$  and  $Y=r*\sin(\theta)$ .



**Figure 6.** Profile views of the observed logits on the polar coordinate system.

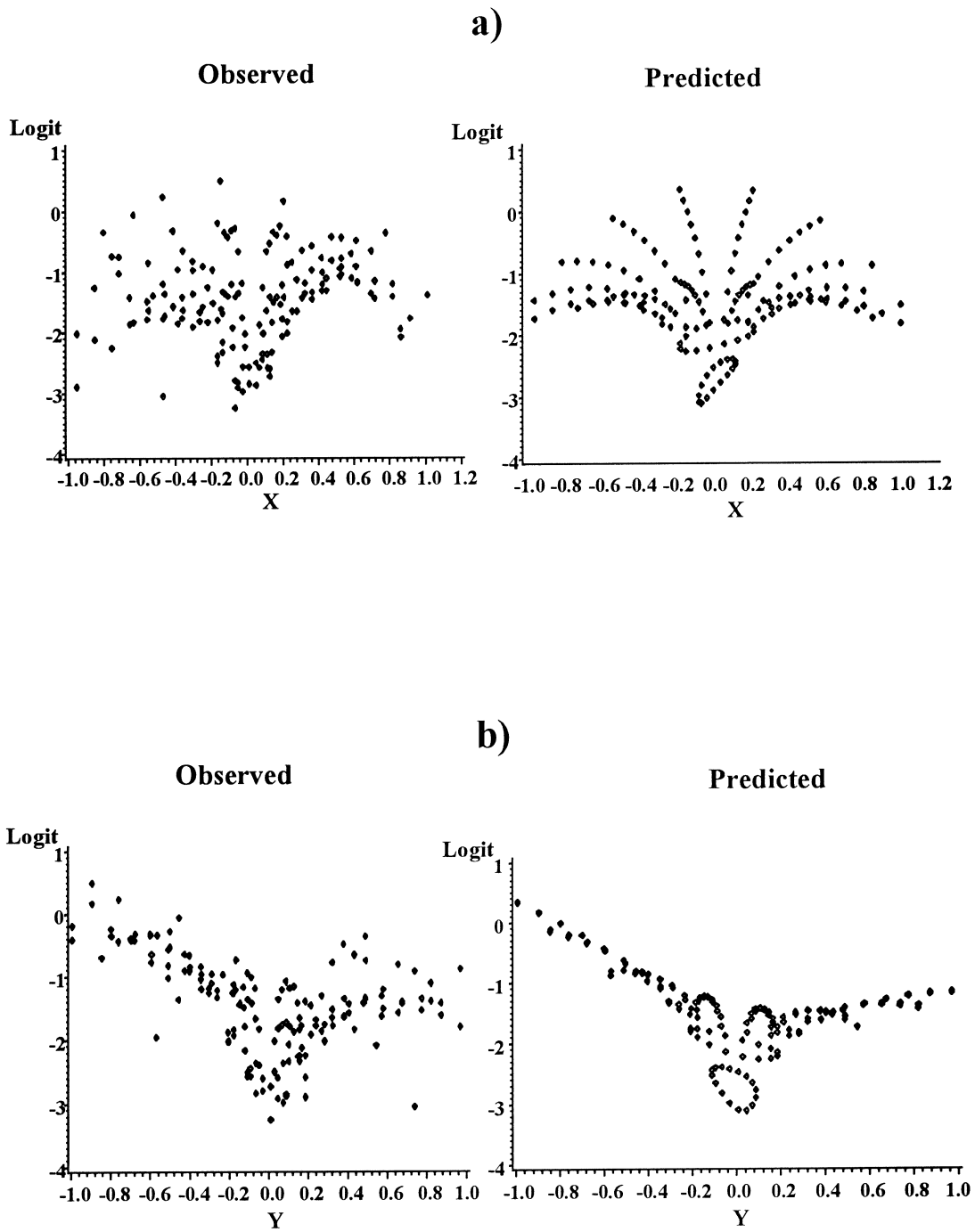
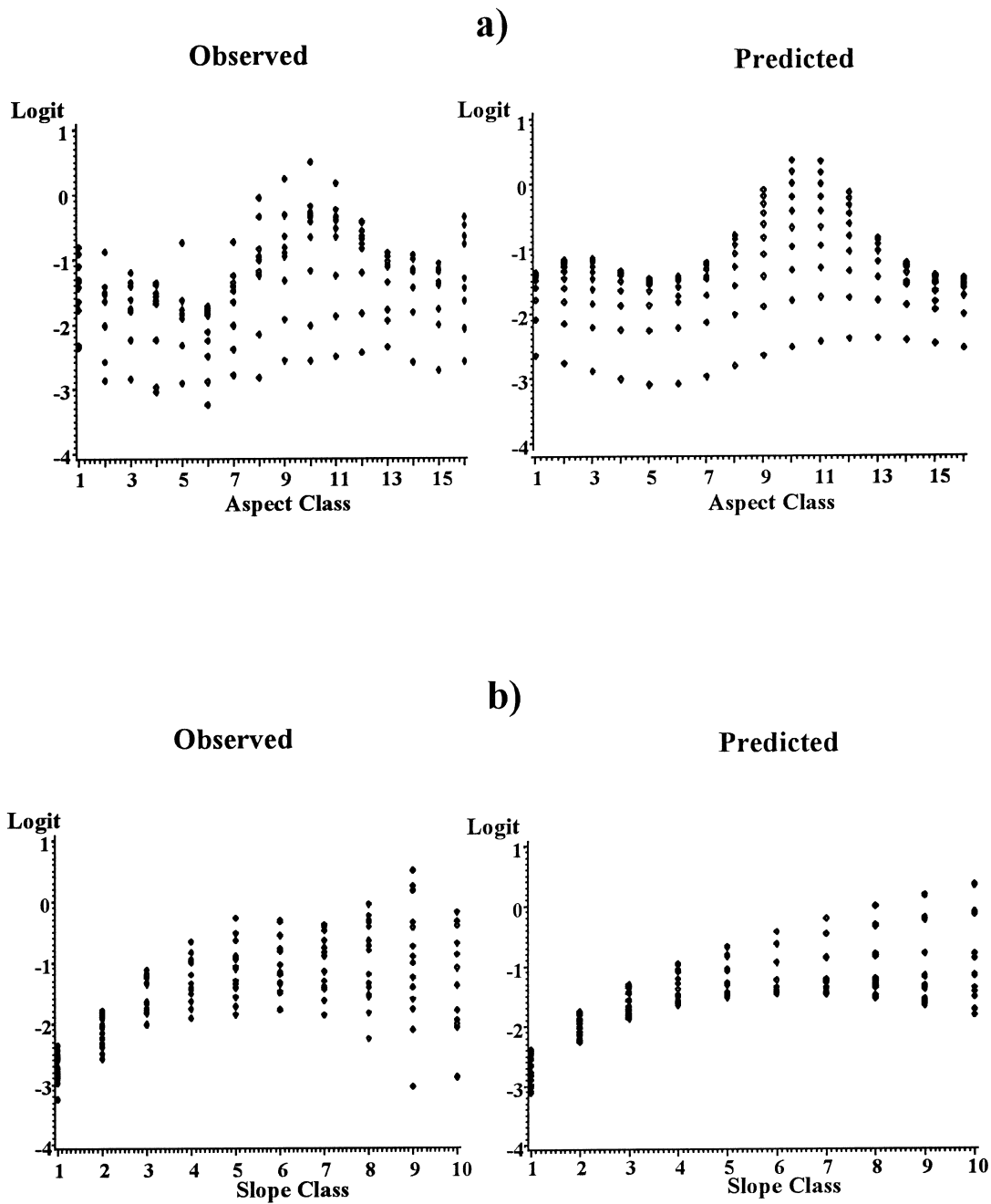
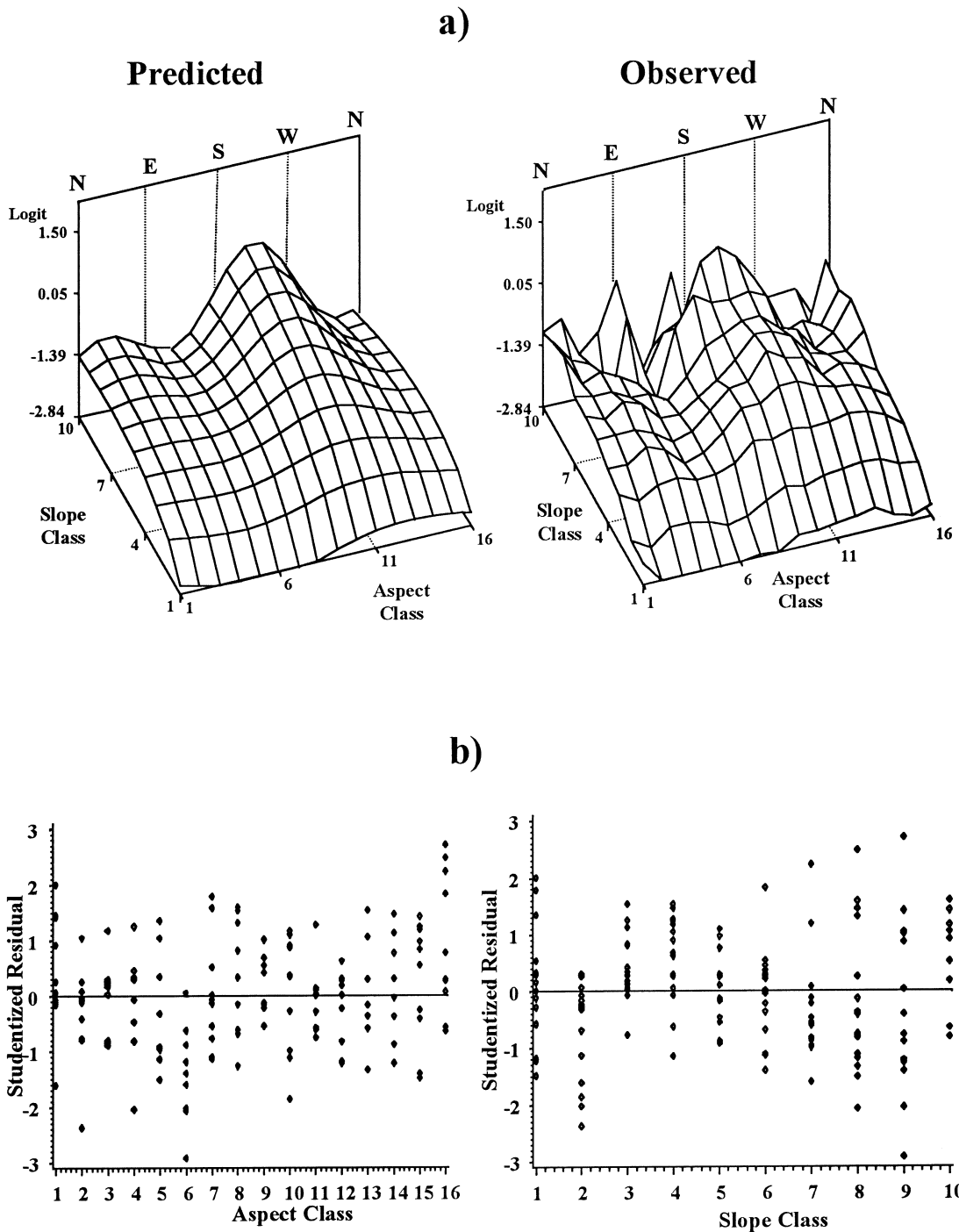


Figure 7. Comparison of observed and predicted logits versus polar coordinate axes X (a) and Y (b).

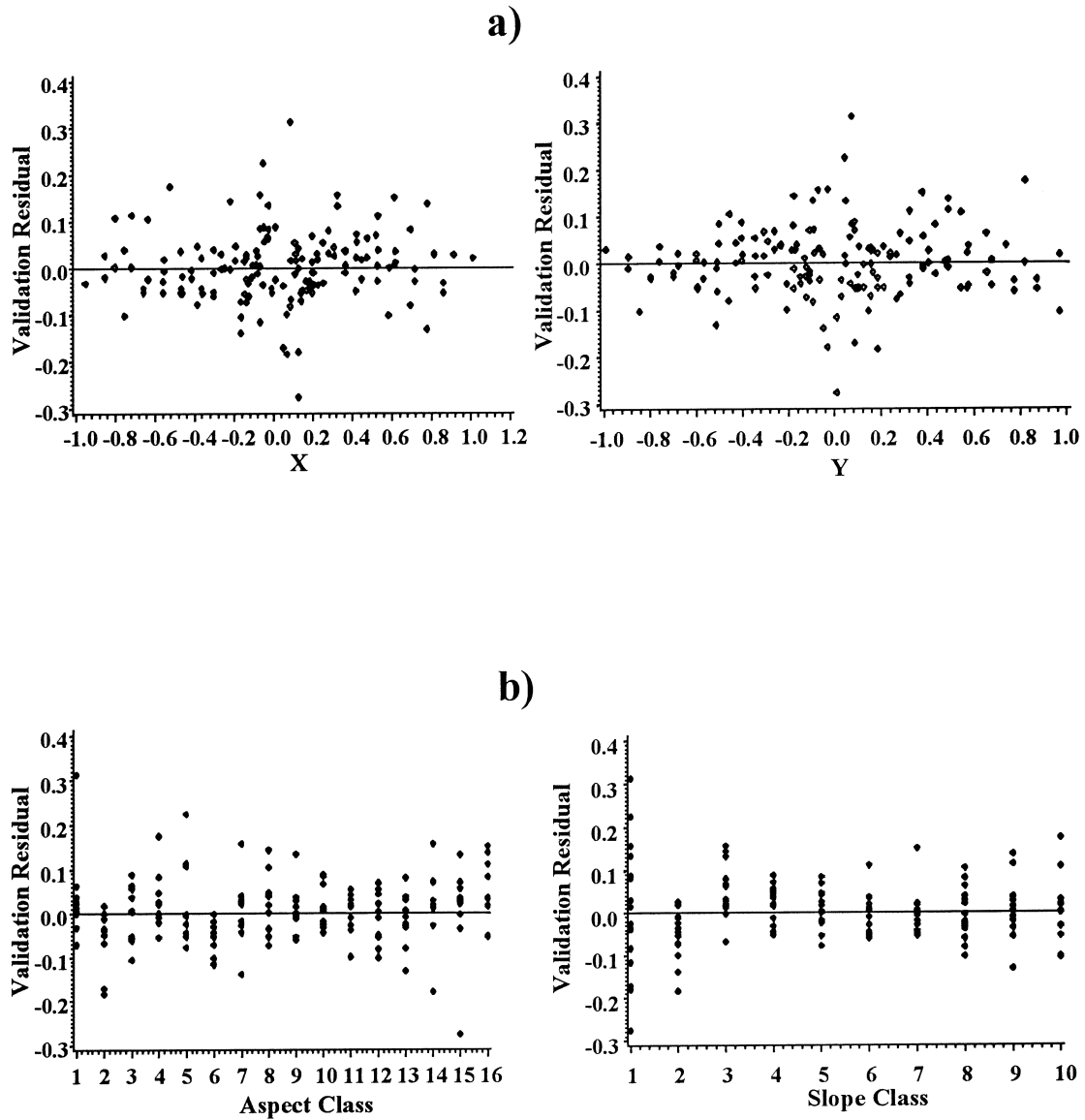




**Figure 8.** Comparison of observed and predicted logits versus aspect (a) and slope (b) classes.



**Figure 9.** Predicted and observed surfaces for Model IV (a) along with the underlying residual structure (b).



**Figure 10.** Validation residuals plotted against polar coordinates (a) and slope and aspect (b).