Common Loon Habitat Modeling in Northern Lower Michigan Using Binary Logistic Regression

Maxwell A. Field
Department of Biology, Central Michigan University

Abstract

Common Loon (*Gavia immer*) is listed as a threatened species in Michigan. It is estimated that only near 250 nesting pairs present throughout the state and the Loon population is declining. In the summer of 2007, about 50 inland lakes and approximately 120 Loons are surveyed in Charlevoix, Cheboygan and Emmet counties of Northern Lower Michigan. In this paper, logistic regression models are used to analyze Loon habitats characteristics to understand and model Loon presence/absence and nesting activity. The best logistic regression models are obtained by the statistic software, MINITAB. Analytical results are given which are helpful for conservation biologist to identify critical habitat for Common Loons in Michigan.

I. Introduction

Common Loon (*Gavia immer*) is listed as a threatened species in Michigan. It is estimated that only near 250 nesting pairs present throughout the state and the Loon population is declining (Williams, pers. comm.). To understand the Loon population distribution, a field survey was conducted in the summer of 2007 for the three counties in the Northern Lower Michigan: Charlevoix, Cheboygan and Emmet (Figure 1). About 50 inland lakes and approximately 120 Loons are surveyed. Specifically, the survey was to gather data that indicates whether an inland lake is used for nesting, feeding, or absent of Loon activity. The field data make it possible to explore characteristics of suitable Loon habitats and to eventually predict Loon presence/absence and nesting activity. The literature of Loon research suggests that there are an extremely large number of variables that could possibly predict this information. Most of variables can be grouped into five categories: (1) Direct Human Disturbance Factors; (2) Indirect Human Disturbance Factors; (3) Abiotic Inland Lake Level Factors; (4) Biotic Inland Lake Level Factors; and (5) Competitive Factors between Loons and other organisms that live on or around inland lakes.

The initial research question is that whether more Loon nesting habitats, such as natural shoreline vegetation types, need to be protected in order to help the Loon population start increasing again. It is anticipated that analytic and modeling efforts can act as an assistant/management tool that can help answer the question and determine which of the following is true about Loon habitats in Michigan. For example, there may be enough nesting habitat, just too much human disturbance associated with that particular lake. It could be even simpler, where there are not enough Loons to claim the available nesting habitat. If the existing Loon presence cannot be successfully predict with the collected physical variables, additional variables, such as toxicants and/or pollutants or other abiotic factors are causing this decreasing trend (mercury, lead, etc.).

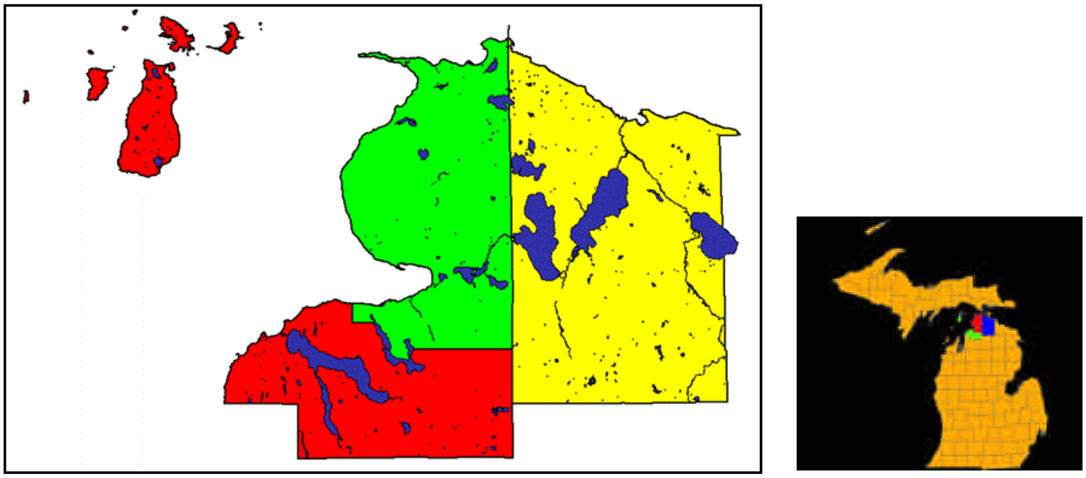

Figure 1. Study Area: Charlevoix County (shaded in red), Cheboygan County (shaded in yellow) and Emmet County (shaded in green).

II. Background

Most of the existing common Loon literature involves some forms of individual Loon level studies or inland lake level water quality research, such as mercury testing.

There are only a few efforts for Loon habitat modeling. For example, Earnst et al. (2006) propose a binary multivariate logistic regression model that predicts Yellow-Billed Loon (*Gavia adasmii*) presence/absence in Alaska. The model is develped based on survey data of 757 lakes. Variables gathered for this model include: (a) natural log lake area; (b) maximum lake depth, classified as shallow (<1.6 m), medium (1.6–4 m), or deep (>4.0 m); (c) proportion of the shoreline in aquatic vegetation; (d) an index of shoreline complexity; (e) hydrological connectivity, classified as connected if the lake's perimeter was within 100 m of a stream; (f) distance from the lake's centroid to a river and (g) distance from the lake's centroid to the coast of the Beaufort Sea. All of these parameters are explanatory except the shoreline complexity index, the ratio of the lake's perimeter to the perimeter of a circle of equal area (Earnst et al. 2006). This provided a unitless index with a minimum of one, which is equivalent to the shoreline complexity of a circle. Lake area, depth, proportion of shoreline in aquatic vegetation, shoreline complexity, hydrological connectivity (stream present within 100 m or absent), and an area–connectivity interaction are found significant predictors of yellow-billed Loon presence in a multivariate logistic regression model while distance to nearest river or Beaufort Sea coast is irrelevant (Earnst et al. 2006).

Nacci et al. (2005) present a binary logistic regression model that predicts risks to wildlife populations in New Hampshire. The authors use field-gathered survey data and GIS data to estimate potential Mercury contamination to Loons. This Loon habitat suitability model is developed using a subset of monitored Loon presence/absence field data at 586 lakes using GIS data. Variables they used in their model include: (a) Minimum Distance to a Loon Lake; (b) Lake Elevation; (c) Ratio (lake perimeter/sqrt.(area)); (d) Perimeter (excluding islands); (e) Lake Area (excluding islands); and (f) Number of Islands/Lake. Initial results for a local habitat model suggest that Loons tend to occupy larger, deeper lakes with islands and/or extensively convoluted shorelines, and located near lakes occupied by other Loons. When compared to the subset of lakes left out during creation of the model, the logistic regression model is successful at predicting 80% of the confirmed Loon presence/absence lakes.

III. Method

A. Predictor and response variables

Potential explanatory variables examined in this paper are either gathered in the field or downloaded online from Michigan Center for Geographic Information. Field collected variables include: Maximum Lake Depth (Feet), the number of Islands, Water Recreation Activity Index, the number of Loons observed, the number of Loon nests observed, and the number of Loon chicks observed. These Loon count variables are put together to form the dependent variables: the Loon presence/absence data and the Loon nest presence/absence data. The variables collected and/or verified with GIS include: Inland Lake Area (Hectares), Inland Lake Perimeter (Meters), the number of Islands, Shoreline Complexity (Perimeter/Sqrt(Area)), Percent Shoreline Development (Human Structures), Percent Shoreline Wetland (Aquatic Vegetation), Water Recreational Activity Index (1 = little to no activity, 2 = average boating activity, 3 = maximum boating activity), and Minimum Distance to a Loon Lake with verified Loon Presence. Two different logistic regression models are constructed to understand the overall Loon presence/absence (regardless of feeding or nesting activity) and Loon nest presence/absence (based on either chicks or nests observed). All the variables are summarized in Tables 1 and 2.

Table 1. Logistic Regression Predictor Variables

Predictor Variable	**Source**	**Type**
Area (Hectares)	GIS	Continuous
lnArea (Hectares)	Natural Log Transformed	Continuous
Max. Lake Depth (Feet)	GIS	Continuous
Perimeter (Meters)	GIS	Continuous
# Islands (Count)	GIS / Loon Surveys	Continuous
Perimeter-Area (Ratio)	GIS	Continuous
% Shoreline Development	GIS	Continuous
% Shoreline Wetland	GIS	Continuous
Min. Dist. to Loon Lake (Meters)	GIS	Continuous
Water Recreation Index (1, 2. 3)	Loon Surveys	Categorical

Table 2. Logistic Regression Response Variables

Response Variable	**Binary Code**
Overall Loon Presence/Absence (Model 1)	Y = 1 (Presence) or Y = 0 (Absence)
Loon Nest Presence/Absence (Model 2)	Y = 1 (Presence) or Y = 0 (Absence)

IV. Results

A. Best Subsets selection

The Best Subsets selection method of the statistic software MINITAB is used to determine which of the potential predictor variables to include in the final models. The Best Subsets model selection method is chosen over the Forward or Backward Stepwise model selection method because Stepwise is more useful when dealing with an extremely large number of predictor variables. In this case, a maximum number of variables is 10.

For the first model (overall Loon presence/absence), best subsets regression results are as follows (Table 3). The top two models suggested by MINITAB are highlighted with the highest adjusted R^2: 28.4 and 25.4 respectively. The two models are represented as Model 1_1 and Model 1_2, which are further compared. The logistic regression results of Model 1_1 and Model 1_2 are showed in Tables 4a and 4b. For the second model (Loon nest presence/absence), best subsets regression results are as follows (Table 5). The top two models suggested by MINITAB are highlighted with the highest adjusted R^2: 30.7 and 29.8 respectively. The two models are represented as Model 2_1 and Model 2_2, which are further compared. The logistic regression results of Model 2_1 and Model 2_2 are showed in Tables 6a and 6b.

Table 3. Results of best subsets regression for model #1 (overall Loon presence/absence)

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	lnArea	perimeter	Ratio	Max_Depth	Islands	MinDist_LL	%Wet	%Shline
1	14.8	12.7	6.6	0.43714					X			
1	4.4	2.0	12.1	0.46315	X							
2	20.5	16.4	5.7	0.42785	X	X						
2	19.2	15.1	6.3	0.43112					X		X	
3	31.7	26.3	1.8	0.40175	X	X			X			
3	24.6	18.6	5.5	0.42214	X	X				X		
4	35.3	28.4	1.9	0.39605	X	X			X	X		
4	32.7	25.4	3.3	0.40419	X	X		X	X			
5	36.5	27.7	3.2	0.39785	X	X		X	X	X		
5	35.6	26.6	3.7	0.40084	X	X	X		X	X		
6	36.8	26.0	5.1	0.40261	X	X	X	X	X	X		
6	36.7	25.8	5.2	0.40306	X	X		X	X	X		X
7	36.9	24.0	7.0	0.40804	X	X	X	X	X	X		X
7	36.8	23.8	7.1	0.40846	X	X	X	X	X	X	X	
8	37.0	21.7	9.0	0.41401	X	X	X	X	X	X	X	X

Table 4a. Results of logistic regression for model 1_1

```
Variable  Value  Count
Loon_P_A  1         29  (Event)
          0         13
          Total     42
Logistic Regression Table
                                                             Odds     95% CI
Predictor         Coef     SE Coef      Z      P  Ratio  Lower  Upper
Constant      -2.67901     1.53032  -1.75  0.080
lnArea         1.27357    0.547022   2.33  0.020   3.57   1.22  10.44
perimeter   -0.0002195   0.0001085  -2.02  0.043   1.00   1.00   1.00
Islands        1.19013    0.594390   2.00  0.045   3.29   1.03  10.54
MinDist_LL   -0.182502    0.111504  -1.64  0.102   0.83   0.67   1.04

Log-Likelihood = -17.149
Test that all slopes are zero: G = 17.674, DF = 4, P-Value = 0.001
Goodness-of-Fit Tests
Method                Chi-Square  DF      P
Pearson                  37.9484  37  0.426
Deviance                 34.2981  37  0.596
Hosmer-Lemeshow           7.0381   8  0.533
Measures of Association:
(Between the Response Variable and Predicted Probabilities)
Pairs        Number  Percent  Summary Measures
Concordant      310     82.2  Somers' D                  0.65
Discordant       65     17.2  Goodman-Kruskal Gamma      0.65
Ties              2      0.5  Kendall's Tau-a            0.28
Total           377    100.0
```

Table 4b. Results of logistic regression for model 1_2

```
Variable  Value  Count
Loon_P_A  1         29  (Event)
          0         13
          Total     42
Logistic Regression Table
                                                  Odds     95% CI
Predictor         Coef   SE Coef      Z      P  Ratio  Lower  Upper
Constant      -2.75141   1.53375  -1.79  0.073
lnArea         1.23185  0.541745   2.27  0.023   3.43   1.19   9.91
perimeter   -0.0002309 0.0001097  -2.10  0.035   1.00   1.00   1.00
Max_Depth    0.0103548 0.0199097   0.52  0.603   1.01   0.97   1.05
Islands        1.21773  0.598398   2.03  0.042   3.38   1.05  10.92
MinDist_LL   -0.182873  0.111539  -1.64  0.101   0.83   0.67   1.04

Log-Likelihood = -16.999
Test that all slopes are zero: G = 17.975, DF = 5, P-Value = 0.003
Goodness-of-Fit Tests
Method           Chi-Square  DF      P
Pearson             36.0940  36  0.464
Deviance            33.9978  36  0.564
Hosmer-Lemeshow      4.4471   8  0.815
Measures of Association:
(Between the Response Variable and Predicted Probabilities)
Pairs       Number  Percent  Summary Measures
Concordant     313     83.0  Somers' D              0.67
Discordant      62     16.4  Goodman-Kruskal Gamma  0.67
Ties             2      0.5  Kendall's Tau-a        0.29
Total          377    100.0
```

Table 5. Results of best subsets regression for model #2 (Loon nest presence/absence)

```
----------------------------------------------------------------
                                                   M
                                           p   M   i
                                           e   a   n   %
                                           r   x I D
                                         l i   _ s i   S
                                         n m R D l s % h
                                         A e a e a t   l
                                         r t t p n _ W i
                       Mallows           e e i t d L e n
Vars  R-Sq  R-Sq(adj)    C-p         S  a r o h s L t e
----------------------------------------------------------------

   1  32.0       30.3     -2.7  0.41020          X
   1   6.8        4.5     10.4  0.48033      X
   2  34.1       30.7     -1.8  0.40913  X       X
   2  33.2       29.7     -1.3  0.41204          X X
   3  35.6       30.5     -0.6  0.40963  X X     X
   3  34.9       29.8     -0.2  0.41191  X       X X
   4  36.1       29.1      1.2  0.41374  X X     X X
   4  36.0       29.1      1.2  0.41395  X X     X   X
   5  36.4       27.5      3.0  0.41837  X X     X X X
   5  36.1       27.2      3.2  0.41924  X X X   X X
   6  36.4       25.5      5.0  0.42417  X X X   X X X
   6  36.4       25.5      5.0  0.42426  X X     X X X X
   7  36.4       23.3      7.0  0.43032  X X X   X X X X
   7  36.4       23.3      7.0  0.43036  X X X X X X X
   8  36.4       21.0      9.0  0.43679  X X X X X X X X
----------------------------------------------------------------
```

Table 6a. Results of logistic regression for model 2_1

```
Variable   Value  Count
Loon_Nest  1         16  (Event)
           0         26
           Total     42
Logistic Regression Table
                                                Odds     95% CI
Predictor        Coef   SE Coef      Z      P  Ratio  Lower  Upper
Constant     -2.93658   1.26531  -2.32  0.020
lnArea       0.248461  0.234344   1.06  0.289   1.28   0.81   2.03
Islands       1.18774  0.373235   3.18  0.001   3.28   1.58   6.82
MinDist_LL  0.0646730  0.102821   0.63  0.529   1.07   0.87   1.30

Log-Likelihood = -19.978
Test that all slopes are zero: G = 15.864, DF = 3, P-Value = 0.001
Goodness-of-Fit Tests
Method           Chi-Square  DF      P
Pearson             41.7000  38  0.313
Deviance            39.9564  38  0.383
Hosmer-Lemeshow      8.2327   8  0.411
Measures of Association:
(Between the Response Variable and Predicted Probabilities)
Pairs       Number  Percent  Summary Measures
Concordant     348     83.7  Somers' D                 0.68
Discordant      66     15.9  Goodman-Kruskal Gamma     0.68
Ties             2      0.5  Kendall's Tau-a           0.33
Total          416    100.0
```

Table 6b. Results of logistic regression for model 2_2

```
Variable   Value  Count
Loon_Nest  1         16  (Event)
           0         26
           Total     42
Logistic Regression Table
                                              Odds     95% CI
Predictor      Coef   SE Coef      Z      P  Ratio  Lower  Upper
Constant   -2.72464   1.19710  -2.28  0.023
lnArea     0.257638  0.230340   1.12  0.263   1.29   0.82   2.03
Islands     1.18569  0.375957   3.15  0.002   3.27   1.57   6.84

Log-Likelihood = -20.188
Test that all slopes are zero: G = 15.444, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests
Method           Chi-Square  DF      P
Pearson             40.9659  39  0.384
Deviance            40.3767  39  0.409
Hosmer-Lemeshow     10.5731   8  0.227
Measures of Association:
(Between the Response Variable and Predicted Probabilities)
Pairs       Number  Percent  Summary Measures
Concordant     349     83.9  Somers' D                 0.69
Discordant      63     15.1  Goodman-Kruskal Gamma     0.69
Ties             4      1.0  Kendall's Tau-a           0.33
Total          416    100.0
```

B. Evaluation of Model – Final Diagnostics

The MINITAB outputs of logistic regressions are examined: p-values, odds ratios, three goodness of fit tests, concordant pairs, and three summary measures (Tables 4a, 4b, 6a and 6b). The relationships between predictor variables are the focus of the analysis rather than the overall predictive power of the models. Thus, the significance level is set to 0.1, that is, any p-value less than 0.1 is considered significant with 90% confidence. All four models 1_1, 1_2, 2_1, and 2_2 (the first two for Loon Presence/Absence, and the other two for Loon Nest Presence/Absence) have at least one significant predictor with a p-value less than 0.05. These variables also had high odds ratios (usually >1), indicating their high predicting power within the models.

The p-values from the three goodness of fit tests are all greater than 0.05 in the four model. This indicates that the data fit the model relatively well. Hosmer-Lemeshow is always the smallest of the goodness of fit tests. This could be explained by the fact that Hosmer-Lemeshow specifically is looking for repeat values of predictors, such as the number of Islands, Shoreline Development, Max. Depth, and Minimum Distance to a Loon Lake especially. Concordant pairs are usually >80% once again indicating that the data fit the model relatively well. Summary measures are the only indicator that the data wasn't that well of a fit. The Somers' D and Goodman-Kruskal Gamma measures are usually greater than 0.50 but less than 0.75 indicating a weaker fit of the data to the model. The last summary measure, Kendall's Tau-a always is less than 0.50 indicating a poor fit of the data to the model.

Three logistic regression diagnostic graphics are created for each of the four models: Delta Chi-Square vs. Probability, Deviance vs. Cases, and Residuals (EPR) vs. Standardized Pearson (SPRE) (Figures 2 and 3). Figures 2 shows that Model 1_1 has outliers that are less extreme compared to Model 1_2 (see Y-axis). The Delta Chi-Square vs. Probability graphs for Model 1_1 appear to be a better fit since the outlier in the middle of the graph is closer to the general trend of Model 1_1. While examining the residuals, a Lowess Smoothing factor is added. Because the smoothing factor stays closer to zero along the Y-axis for Model 1_1, it is derived that Model 1_1 is a better fit overall than Model 1_2 although there are outliers for both Models. Thus, Model 1_1 is accepted while Model 1_2 is rejected.

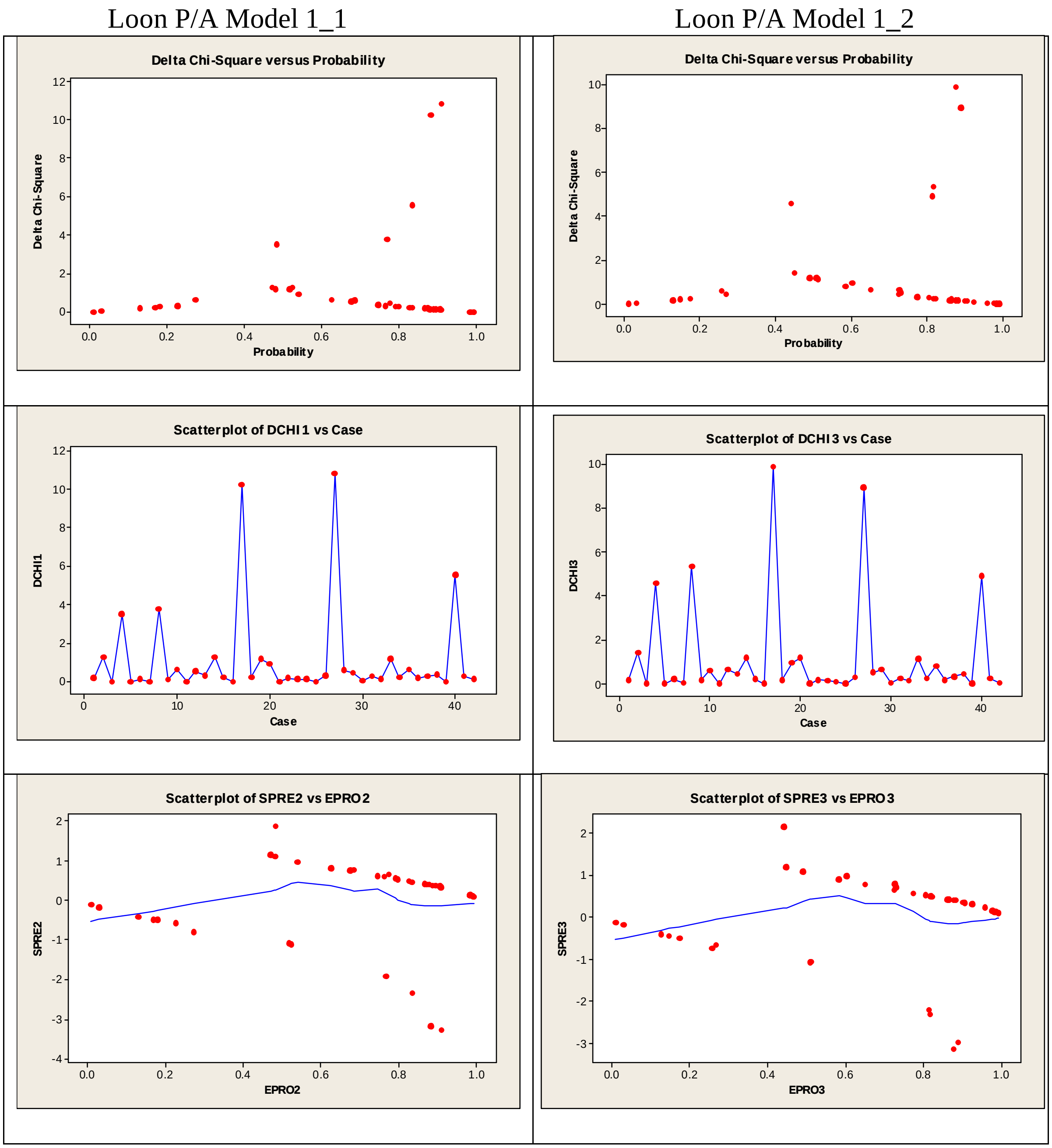

Figure 2. Logistic regression diagnostic graphs of Models 1_1 and 1_2 for overall Loon Presence/Absence

Loon Nest P/A Model 2_1 Loon Nest P/A Model 2_2

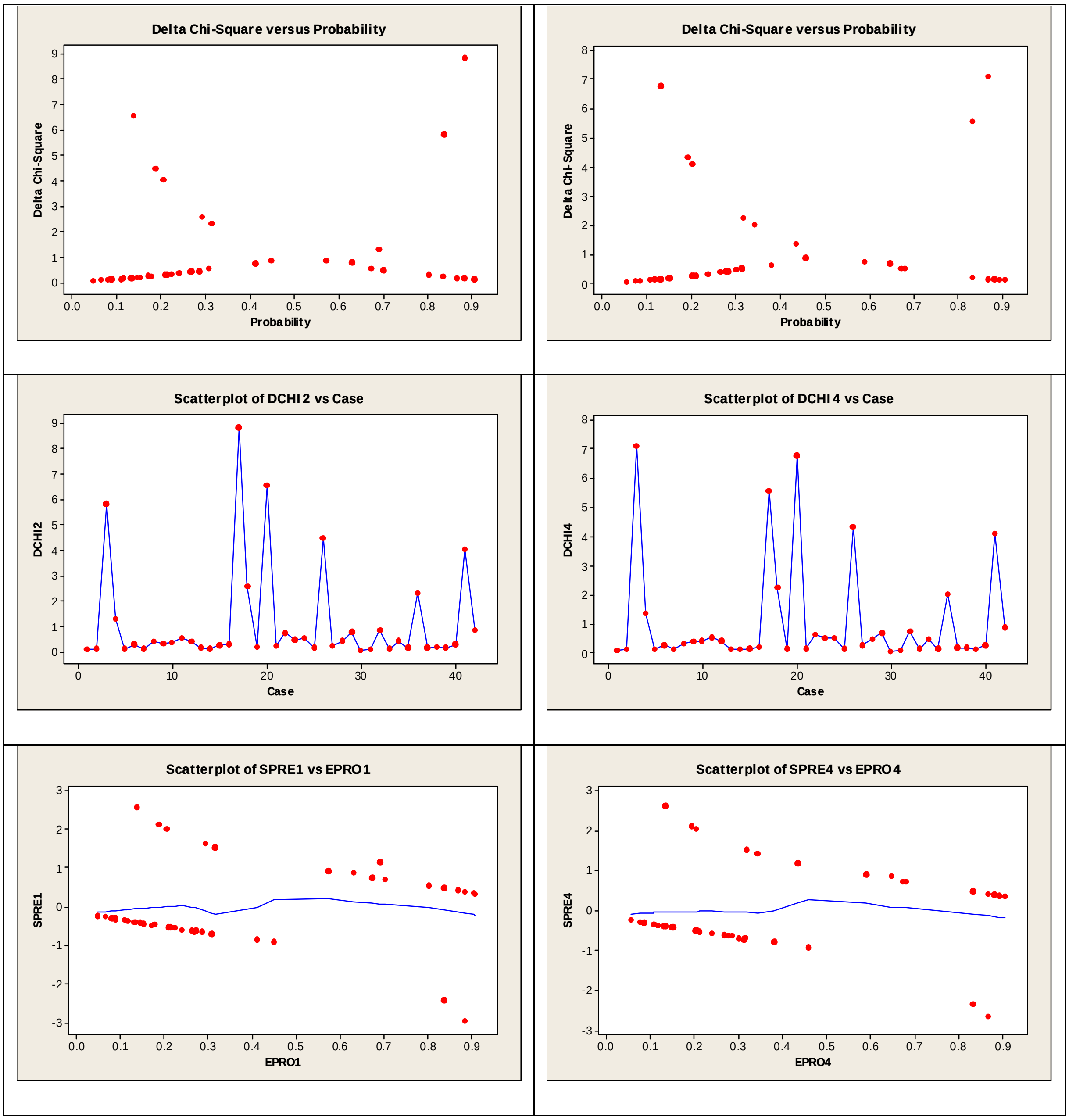

Figure 3. Logistic regression diagnostic graphs of Models 2_1 and 2_2 for Loon nest Presence/Absence

Figure 3 shows that both Models 2_1 and 2_2 have the same number of outliers. The Delta Chi-Square vs. Probability graphs for Model 2_2 appears to be a better fit since the outlier on the middle-right of Model 2_1 doesn't exist for Model 2_2. This outlier deviates from the connected "scissor" pattern. The same parameters are used to create the Lowess smoothing factor from Models 1_1 and 1_2. Because the smoothing factor stays closer to zero along the Y-axis for Model 2_2, it indicates a better fit than Model 2_1. Although there are outliers for both Models, Model 2_2 seems to be the better fit overall when examining the visual diagnostics and it is accepted.

V. Discussion and Conclusions

A. Summarize Final Model

Model #1 for Overall Loon Presence/Absence:

$Y_1 = -2.67901 + 1.27357X_1 - 0.0002195X_2 + 1.19013X_3 - 0.182502X_4$

where Y_1 is the presence/absence of overall Loons; X_1 is the natural log of lake area; X_2 is the lake perimeter; X_3 is the number of islands and X_4 is the minimum distance to the lake with Loon presence.

Model #2 for Loon Nest Presence/Absence:

$Y_2 = -2.72464 + 0.257638X_1 + 1.18569X_3$

where Y_2 is the presence/absence of Loon nest.

Models #1 and # 2 share the same significant predictors: X_1 and X_3. These two predictors made sense, especially when compared to other research that has been conducted in order to build Loon habitat models (Michigan Loon Recovery Committee). Perimeter was included as a significant predictor in my overall Loon presence/absence model, but it was slightly negative, and seemed to disagree with the correlation between Area and Perimeter.

B. Conclusions and Insight

The research objectives are to build logistic regression models that predict overall Loon presence/absence and the presence/absence of Loon nesting activity. The project sheds light on the relationships between Loon habitat variables that could be used to eventually predict Loon distribution and specifically for nesting activity. First, the initial hypotheses that Human Disturbances would adversely affect Loons are disproved through the analysis of these models. It was expected that increased shoreline developments and human activity would have negatively affected Loon presence regardless of nesting activity. However, neither of the human-related predictors (% Shoreline Development, and Water Recreational Activity Index) are even used in the logistic models because they are not statistically significant. Actually there is a slightly positive correlation between % Shoreline Development and Loon Presence! Though this is against common sense, it is reasonable that Loons and humans could possibly be utilizing similar lakes in terms of living. Humans are more likely to live on inland lakes that are larger, deeper, and potentially have more fish in them. Meanwhile Loons are looking for the same factors in many instances. This is possible because there are still amounts of natural shoreline and/or islands to set up nesting sites, and larger lakes have the potential to support multiple nesting pairs of Loons due to their increased area and potential fish resources for food. It seems that management strategies can be used to promote harmonious co-existence between Loons and human beings.

Secondly, the number of Islands is always a positive significant predictor of overall Loon presence, and Loon nesting presence. This makes sense because Loons prefer to nest on islands so that the eggs are less likely to be eaten by a terrestrial predator (McIntyre 1988). In addition, lakes with multiple islands can act as cover for Loons to hide from human presences. The Michigan Loon Recovery Plan actually found, based on 1980's data, that the presence of islands on an inland lake were significant predictors of Loon nesting activity.

The natural log of lake area is a positive significant predictor of Loon presence for the overall Loon presence/absence model. This could be correlated to the fact that Loons ultimately can be found on all sizes of lakes, small to large. There is a minimum threshold for lakes that cannot be used by Loons around four to six hectares. Loons need at least 40 yards of linear "take-off" space because they are large bodied waterbirds with legs near the rear end for diving. Inland lakes that are below this minimum threshold are believed to not have enough room for this take-off procedure (McIntyre 1988). There are two lakes below the minimum threshold and they are confirmed to have a Loon absence.

Perimeter is a negative significant predictor of Loon presence for the overall Loon presence/absence model. Although the coefficient was very small (-0.0002309), it still came out as significantly negative. This is a confusing significant predictor variable because lakes with a larger perimeter are correlated to lakes having a larger area. Based on that correlation, I would have expected perimeter to be a positive significant predictor.

A limitation of this research is that a relatively small sample size is used (42 lakes). The regression models in the literature of Loon studies have sample sizes of well over 500 inland lakes. Though the low sample size takes away from predicting power of the models, but at the same time still they provide insights on the relationship and importance of selected inland lake attributes that can be used for the future research efforts. More field surveys will be conducted in the summer of 2008. Accessing to the Michigan Loonwatch database will further facilitate better understanding Loon presence/absence and nesting activities.

References

Earnst, S. L., Platee, R., and Bond, L.. 2006. A landscape-scale model of yellow-billed Loon (*Gavia adamsii*) habitat preferences in northern Alaska. Hydrobiologia, (567): 227-236.

McIntyre, J.M.. 1988. The Common Loon: Spirit of Northern Lakes. 1st edition. University of Minnesota Press. Minneapolis, Minnesota.

Michigan Loon Recovery Committee. 1992. A Plan For Recovery of the Common Loon in Michigan. Michigan Loon Preservation Association.

Nacci, D., Pelletier, M., Lake, J., Bennett, R., Nichols, J., Haebler, R., Grear, J., Kuhn, A. Copeland, J., Nicholson, M., Walters, S., and Munns Jr, W.R.. 2005. An Approach to Predict Risks to Wildlife Populations from Mercury and Other Stressors. Ecotoxicology, (14): 283-293.

Williams, Joanne. 2006. Michigan Loonwatch. Michigan Loon Preservation Association. URL: http://www.michiganLoons.org.