Wetlands Evaluation in Ontario: Models for Predicting Wetland Score

NEST Technical Report TR-025
October 1995

by
S. Chisholm
J.C. Davies
G. Mulamoottil
D. Capatos

Ontario

Wetland Evaluation in Ontario: Models for Predicting Wetland Score

NEST Technical Report TR-025
October 1995

by
S. Chisholm,
J.C. Davies,
G. Mulamoottil,
and D. Capatos

NEST Technical Report TR-025

50800
(1.5 k Rp., 30-Oct-95)
ISBN 0-7778-4721-3
ISSN 1192-8182

Copies of this publication are available from the Ontario Ministry of Natural Resources through:

Northeast Science & Technology

60 Wilson Avenue, 2nd Floor
Timmins, Ontario
P4N2S7

or
P.O. Box 3000
140 4th Avenue
Cochrane, Ontario
P0L 1C0

Cette publication spécialisée n'est disponible qu'en anglais

This publication should be cited as:

Chisholm, S., Davies, J.C., Mulamoottil, G., Capatos, D., 1995. Wetland Evaluation in Ontario: Models for Predicting Wetland Score. OMNR. Northeast Science & Technology. TR-025 43p.

Wetland evaluation systems have been developed for northern and southern Ontario by the Ontario Ministry of Natural Resources. The procedure involves ascribing points to over 50 wetland attributes which are divided into four components: Biological, Social, Hydrological and Special Features. Wetlands which achieve a total score of 600 or more points or score 200 or more points in either the Biological or Special Features component are considered to be provincially significant. As the evaluation process is both expensive and time consuming, the development of a technique that could identify those wetlands with the highest potential of being provincially significant would result in considerable savings. Such a technique would be especially valuable in northern Ontario given the large number of unevaluated wetlands that exist. By using the evaluation data from 123 southern Ontario wetlands, this report demonstrates that a small subset of attributes can accurately predict the number of points a wetland would score if it underwent a full evaluation. Using regression analysis, various models were developed to predict a wetland's total score with reasonable accuracy. These models were then modified for application in northern Ontario using the data from 66 wetlands that have been evaluated using the 1993 Ontario Wetland Evaluation System Northern Manual.

Table of Contents

List of Figures

Figure

List of Tables

Table

We would like to thank Neil Maurer, John Parton, Margaret McLaren, and Dean Pheonix for reviewing the penultimate draft of this manuscript and members of the Provincial Wetland Technical Team (WETT) for comments on earlier drafts. Thanks to Louise Clement for proofreading, to Doug Skeggs for co-ordinating the publication of this techical report, and to Diane Wahlman for layout and design.

Wetlands are an important resource which provide significant environmental, economic and social benefits. Despite these values, significant amounts of Ontario's wetland base has been lost to drainage, in fill and land development. It is estimated that since the arrival of European settlers, over three-quarters of the original wetlands in southern Ontario below the Canadian Shield have been lost (Snell, 1987). In response to these ongoing threats, the Ontario government developed the Wetlands Policy Statement under section three of the Planning Act (Ministry of Natural Resources and Ministry of Municipal Affairs, 1992). The goals of this policy statement are to ensure that wetlands are identified and protected through the land use planning process, and to achieve no loss of "provincially significant wetlands".

In order for a wetland to be protected under the policy statement, it must first be evaluated and scored by the Ontario Ministry of Natural Resources (MNR). The evaluation process is explained in the Northern and Southern Manuals of the Ontario Wetland Evaluation System (MNR, 1993a & 1993b). The southern manual applies to the area generally south of the southern edge of the Canadian Shield, while the northern manual applies to the area north of this line.

The evaluation procedure involves assigning points to the different features of a wetland, based on four components: Social, Hydrological, Biological and Special Features. As the score for each component is capped at 250 points, a wetland can score a maximum of 1000 points. Wetlands which achieve a total score of 600 or more points, or score 200 or more points in either the Biological or Special Features components, are considered to be provincially significant.

A limitation of the wetland evaluation system is that it is both costly and time consuming; this presents a serious constraint in northern Ontario where there is a large number of unevaluated wetlands. Consequently, a technique that would enable resource managers to quickly identify high priority wetlands based on their relative importance would allow for the most efficient expenditure of limited evaluation dollars. This approach will also facilitate wetland protection by identifying potentially significant wetlands early on in the planning process. With this information, construction and other development activities in or around these wetlands can be avoided.

OBJECTIVES

The objective of this project is to develop a series of predictive models using the data collected from southern Ontario wetland evaluations. It is expected that these models will be able to demonstrate that a subset of wetland attributes can accurately predict the total score of a wetland and therefore its potential of being provincially significant. Based on the results, the models will then be modified for application in northern Ontario.

METHODS

The data from 115 wetland evaluations that were undertaken in southern Ontario using the Ontario Wetland Evaluation System Southern Manual (MNR, 1993b) were provided by Ecosurveys Limited. This sample is representative of the wetlands that have been evaluated in southern Ontario in terms of site type, geographic location, size, and total score. The wetlands included in this sample have reliable data sets that have been cross-checked. An additional eight wetland evaluations were obtained from the Greater Toronto Area District Office of MNR to supplement these data.

Using this data set of 123 southern Ontario wetland evaluations, an analysis was undertaken to determine whether different combinations of wetland attributes have a significant influence on the total wetland score. The ultimate aim was to develop a series of linear models that could accurately predict the number of points a wetland would score if it underwent a full evaluation. The resulting models were then modified for application in northern Ontario using the data from 66 northern Ontario wetland evaluations provided by MNR Northeast Science & Technology in Cochrane. These evaluations were undertaken using the Ontario Wetland Evaluation System Northern Manual (MNR, 1993a).

Potential variables for analysis were identified from the evaluation manuals and grouped into two categories, Group "A" and Group "B". The values of Group "A" variables can be determined without field work, from sources such as aerial photographs, maps, and MNR District and Area offices. The scoring of Group "B" variables requires visits to the field and they were therefore excluded from the models. A complete listing of the variables considered for inclusion in the models is provided in the Results section of this report.

To determine which factors influence a wetland's total score, correlation and multiple regression analysis were applied to the data. Correlation measures the strength of association between two variables whereas multiple regression analysis describes the degree of association between a single dependent variable and several independent variables. These tests were conducted using SYSTATR statistical software run on a personal computer. The University of Waterloo Statistical consulting service was used extensively to ensure that the testing procedures and interpretations of the data were appropriate.

The purpose of measuring correlation was to ensure that the predictor variables are statistically independent of each other at the 0.05 critical level. Multicollinearity among the predictor variables can inflate the standard errors and threaten computational accuracy of the regression coefficients. Correlation coefficients were calculated using Spearman's Rank Correlation. Under this method, the influence of large values is reduced because the data are converted into ranks prior to calculating the coefficients.

A second assumption of multiple regression is that the residuals follow a normal distribution. To ensure that this condition was met, the skewness of each variable was measured. Those that were significantly skewed, either positively or negatively, were transformed and retested prior to further analysis. In such cases, log transformations were able to correct this condition.

Multiple regression models have a variety of roles including:

i) providing a good description of the behavior of the response variable;
ii) predicting future response and estimation of mean response;
iii) extrapolation or prediction of response outside the range of existing input data;
iv) estimation of parameters;
v) control of a process by varying levels of input; and
vi) developing realistic models of the process (Rawlings, 1988, p. 169).

The regression coefficient measures the percentage of variation in the dependent variable which is explained by variations in the independent variables taken together. Any remaining variation is attributed to the unexplained random error (residual error).

Multiple regression is defined by the following linear model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon$$

where y represents the value of the dependent variable, β_0 is the term of the y-intercept or the value of y when $x = 0$ (referred to as the intercept), $\beta_1 \ldots \beta_n$ are the regression coefficients, $x_1 \ldots x_n$ are the independent variables, and E denotes the residual error.

The resulting models were evaluated according to the adjusted R^2 coefficient which is R^2 "adjusted" for the number of independent variables or degrees of freedom included in the model. The adjusted R^2 is calculated by the following formula:

$$\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{(n-p)}$$

R^2 is the coefficient of determination, n is the number of cases and p is the number of predictors including the constant.

In the models presented in this report, the predicted (or dependent) variable is the total wetland score; the individual wetland attributes are the predictor (or independent) variables. An alternative approach was also undertaken in which separate predictions were made for the scores of the Biological, Social, Hydrological and Special Features components. The predictive accuracy of this method however was considerably weaker due to its poor ability at predicting the score of the Special Features component.

Model development began by deriving a subset of the Group "A" variables which have the greatest influence on the total wetland score. The method used for selecting this subset was interactive backward stepwise regression. This process begins by considering all the candidate variables and removing the least significant predictor in the first step. Stepping continues until no insignificant variables remain in the model. The significance of an independent variable was determined by its T-score which tests the significance of the association between itself and the dependent variable, adjusted for the combined effects of the other variables remaining in the equation. The reported tolerance levels, which measure collinearity amongst the remaining variables were also used as a criterion for variable removal. Since the amount that a predictor variable influences the regression depends on the other variables present in the model, various other combinations of Group "A" variables were examined and the resulting adjusted R^2 values were compared. The purpose of this approach was to ensure that the full range of possible models was considered.

In order to verify that the resulting models are statistically valid, various diagnostic tests were performed on the residuals which are the differences between the observed and predicted values of the dependent variable due to deviations from the fitted model. The purpose of these tests was to ascertain that the errors are normally distributed, have a constant variance and are independent of each other.

The models developed for the southern Ontario data were applied to the northern data and the results were compared. Based on the results, the models were then modified to obtain the highest adjusted R^2 values while retaining as many of the original predictor variables as possible.

This section is divided into two parts. Part A pertains to the models developed for the southern data, while Part B deals with the north.

PART A: SOUTHERN ONTARIO WETLAND ANALYSIS

The variables that were considered for inclusion in the southern Ontario models are listed in Table 1. This table contains the shorthand name of each variable, its full name and section number as it appears in the Southern evaluation manual, and the Group under which it has been categorized. As previously discussed, Group "A" variables are those which can be scored without field work and were the only ones used to develop the models presented. Group "B" variables require visits to the field and were therefore excluded from the models. Table 2 explains the rationale for categorizing the Group "B" variables. For two variables, AVCH (Aboriginal Values & Cultural Heritage) and FISHAB (Fish Habitat), the data provided were incomplete and they were therefore classified under Group "B".

TABLE 1: VARIABLE NAMES

VARIABLE	FULL NAME	SECTION IN South. Manual	GROUP
SIZE1	Actual Wetland Size (ha)		A
1. BIOLOGICAL COMPONENT			
GGD	Growing Degree Days	1.1.1	B
WLTYPE	Wetland Type	1.1.2	A
SITE	Site Type	1.1.3	A
NOTYPE	Number of Wetland Types	1.2.1	A
VEG	Vegetation Communities	1.2.2	B
HABDIV	Diversity of Surrounding Habitat	1.2.3	A
WPROX	Proximity to Other Wetlands	1.2.4	A
INTER	Interspersion	1.2.5	B
OPWAT	Open Water Types	1.2.6	A
SIZE2	Size (Biological)	1.3	B
2. SOCIAL COMPONENT			
EVP	Economically Valuable Products (Total)	2.1	A
REC	Recreational Activities	2.2	A
AESTH	Landscape Aesthetics (Total)	2.3	A
EDTOT	Education and Public Awareness (Total)	2.4	A
HPROX	Proximity to Areas of Human Settlement	2.5	A
OWNER	Ownership	2.6	A
SIZE3	Size (Social)	2.7	B
AVCH	Aboriginal Values and Cultural Heritage	2.8	B
3. HYDROLOGICAL COMPONENT			
FLOOD	Flood Attenuation (Total)	3.1	A
WQI	Water Quality Improvement (Total)	3.2	B
CARBON	Carbon Sink	3.3	B
SEC	Shoreline Erosion Control	3.4	A
TGR	Groundwater Recharge (Total)	3.5	B
4. SPECIAL FEATURES COMPONENT			
RTOT1	Rarity (Wetland Type)	4.1.1	A
RTOT2	Species Rarity (Total)	4.1.2	B
SGFT	Significant Features and Habitats(Total)	4.2	B
FISHAB	Fish Habitat (Total)	4.2.6	B
AGE	Ecosystem Age	4.3	A
TOTAL	Total Wetland Score		

TABLE 2: RATIONALE FOR CATEGORIZING VARIABLES UNDER GROUP "B"

VARIABLE	CATEGORIZATION RATIONALE
GGD	Requires knowledge of soil types
VEG	Requires knowledge of vegetation species
INTER	Requires knowledge of vegetation boundaries
SIZE2	Requires the complete score for Biodiversity sub-component
SIZE3	Requires the complete score for Social component
AVCH	Data incomplete for this variable
WQI	Requires knowledge of vegetation types
CARBON	Requires knowledge of soil types
TGR	Requires knowledge of soil types
RTOT2	Requires knowledge of the endangered and significant species present
SGFT	Requires knowledge of the endangered and significant species present
FISHAB	Data incomplete for this variable

The minimum and maximum values, mean, standard deviation, skewness, and median of each Group "A" variable are listed in Appendix A. The results indicate that the data for SIZE1, HABDIV and AGE are highly skewed and therefore do not represent a normal distribution, based on the standard error test (Wilkinson, 1989). To correct this, these variables were transformed logarithmically and renamed SZ1LOG, HABLOG and AGLOG respectively. Their statistics are also listed in Appendix A. The data reveal that the mean wetland size of the sample is 246.6 ha with a mean total score of 577.7 points.

Through the process of interactive stepwise regression, various combinations of Group "A" variables were tested in order to obtain the highest value possible for the adjusted squared multiple regression (adjusted R^2). Based on the results, two separate series of models were developed (series S1 and S2) which are listed in Table 3. The S1 series contains the highest adjusted R^2 values that were obtained for the set number of independent variables included in each model. These values are listed in the far right column of the table.

The bold, upper-case letters in the left column of Table 3 identify the individual models within the two series. The coefficient values of the independent variables are determined by reading the table horizontally and referring to the variables' names at the top of each section. For example, for model S1B the coefficient values for the variables SZ1LOG and RTOT1 are 32.850 and 1.424 respectively. Each model also includes a constant which, as previously mentioned, is the y-intercept of the regression equation. Appendix B contains the complete statistical data for these models, including the standard errors of the coefficients. The terms that are used in this output are explained in the glossary at the end of this report. The procedure for applying the models to actual data is explained in the following section entitled Model Application Procedure.

TABLE 3: SOUTHERN MODELS (N=123)

S1 MODELS										
Model	Constant	SZ1LOG	RTOT1	REC	FLOOD	OPWAT	HPROX	WPROX	EDTOT	Adj R^2
A	35.462	33.442	1.299	1.693	1.533	3.582	2.141	6.316	1.751	0.817
B	23.297	32.850	1.424	2.002	1.531	3.596	2.581	6.494	-	0.804
C	77.151	33.112	1.461	2.094	1.370	3.476	2.191	-	-	0.789
D	131.674	30.628	1.473	2.181	1.333	3.045	-	-	-	0.770
E	168.414	32.578	1.527	2.194	1.113	-	-	-	-	0.740
F	224.745	41.130	1.733	1.222	-	-	-	-	-	0.646
G	234.384	47.808	1.718	-	-	-	-	-	-	0.593

S2 MODELS										
Model	Constant	EVP	RTOT1	REC	FLOOD	OPWAT	SZ1LOG	EDTOT	HPROX	Adj R^2
A	78.867	2.140	1.346	1.748	1.356	3.249	22.468	1.744	1.815	0.815
B	124.301	2.064	1.327	1.749	1.329	2.917	21.052	2.129	-	0.803
C	124.020	2.139	1.484	2.147	1.312	2.820	19.350	-	-	0.782
D	143.866	4.105	1.430	2.365	1.416	2.818	-	-	-	0.762
E	176.203	4.444	1.482	2.371	1.211	-	-	-	-	0.737
F	249.201	5.460	1.689	1.373	-	-	-	-	-	0.621
G	267.515	6.394	1.662	-	-	-	-	-	-	0.551

In order to maximize flexibility in predicting wetland scores, the S2 model series was developed as the "next best" alternative and is presented in the lower half of Table 3. The models in this series may be useful in cases where it is difficult to verify the presence of or the distance to nearby wetlands since it does not include the variable WPROX. The adjusted R^2 value of these models is slightly lower than their counterparts in the S1 series. For models S2A to S2F, the relative difference is no greater than five percent. The largest variation is between models S2G and S1G at approximately seven percent (0.551 versus 0.593). As these differences are small, the S2 models are considered to be a reliable alternative. The statistical data for these models are also included in Appendix B.

The number of independent variables that can be included in the models is limited by the size of the sample. According to statistical principles, there should be approximately 10 to 15 observations per independent variable. For the southern Ontario models, up to eight independent variables have been incorporated as there are 123 observations (models S1A and S2A). The subsequent models (S1B to S1G, and S2B to S2G) were derived by removing the independent variable that reduced the adjusted R^2 value of the preceding model by the least amount. This process continued until only two variables and the constant remained (models S1G and S2G).

Generally, it is suggested that models S1/S2C to G be utilized. Although models S1/S2A and B report higher adjusted R^2 values, these increases are likely due to the models' improved ability to fit the data set from which they were developed.

This does not necessarily imply that they will be better at predicting the scores of wetlands that are not part of this sample. The difference between the fitting accuracy and the predictive accuracy of regression models is explained in the section Test of Model Accuracy.

As the output in Appendix B indicates, the variability of the data in models S1A and S1B is such that the standard error of the constant is relatively large in relation to its estimated coefficient value. This is evident by the large P(2 TAIL) values of 0.259 and 0.469 respectively which mean that neither constant is significant at the 0.05 critical level. The constants are retained however, since it is not likely that the value of the y-intercepts are equal to zero. That is, if a wetland does not score any points on the independent variables contained in models S1A or S1B, it is highly unlikely that its total score would be zero, since it would score points on the other attributes that are not included. For all of the other variables present in the southern models, the standard errors are small compared to their corresponding coefficient values. As their reported P(2 TAIL) values are either 0.008 or less, these variables are highly significant below the 0.01 critical level. These high significant results are, in part, a function of the stepwise regression procedure employed (Weisberg, 1985 p. 214 -215).

A correlation matrix of Spearman's Rank Correlation coefficients was also calculated between all of the variables included in the models. These matrices are presented in Appendix C. The strongest relationship is between EVP and SZ1LOG at 0.813. Although this value is high, it does not pose a problem since these variables are not present together in any of the same models. The other values are much smaller which indicate that multicollinearity is not a problem with these data unless some form on multicollinearity involving more than two variables is present.

In regression analysis, it is assumed that the residual errors are normally distributed. Using model S1A, the residual errors were plotted against the expected values. The results confirm that this assumption was not violated. This and other diagnostic plots are located and further explained in Appendix D. Draper and Smith (1981) discuss these and other methods of examining regression residuals.

The procedure for applying the regression models presented in this report is outlined in the following steps:

1. Select the model series and model that is most applicable, given the availability of information on the wetland being tested.

2. Determine the number of points that the wetland would score for each variable in the model, according to the procedure explained in the appropriate evaluation manual. For the SZ1LOG variable, determine the size of the wetland in hectares and calculate its natural logarithmic value.

3. Multiply each score by the coefficient for that variable and sum the resulting scores.

4. Add the coefficient of the constant. The resulting total is the model's predicted wetland score.

Example

A 14 hectare wetland scores the following for each attribute listed below:

SZ1LOG = 2.639

RTOT1 = 60
REC = 68
FLOOD = 70
OPWAT = 8
HPROX = 26
WPROX = 8

Using model S1B, the predicted total score for this wetland is:

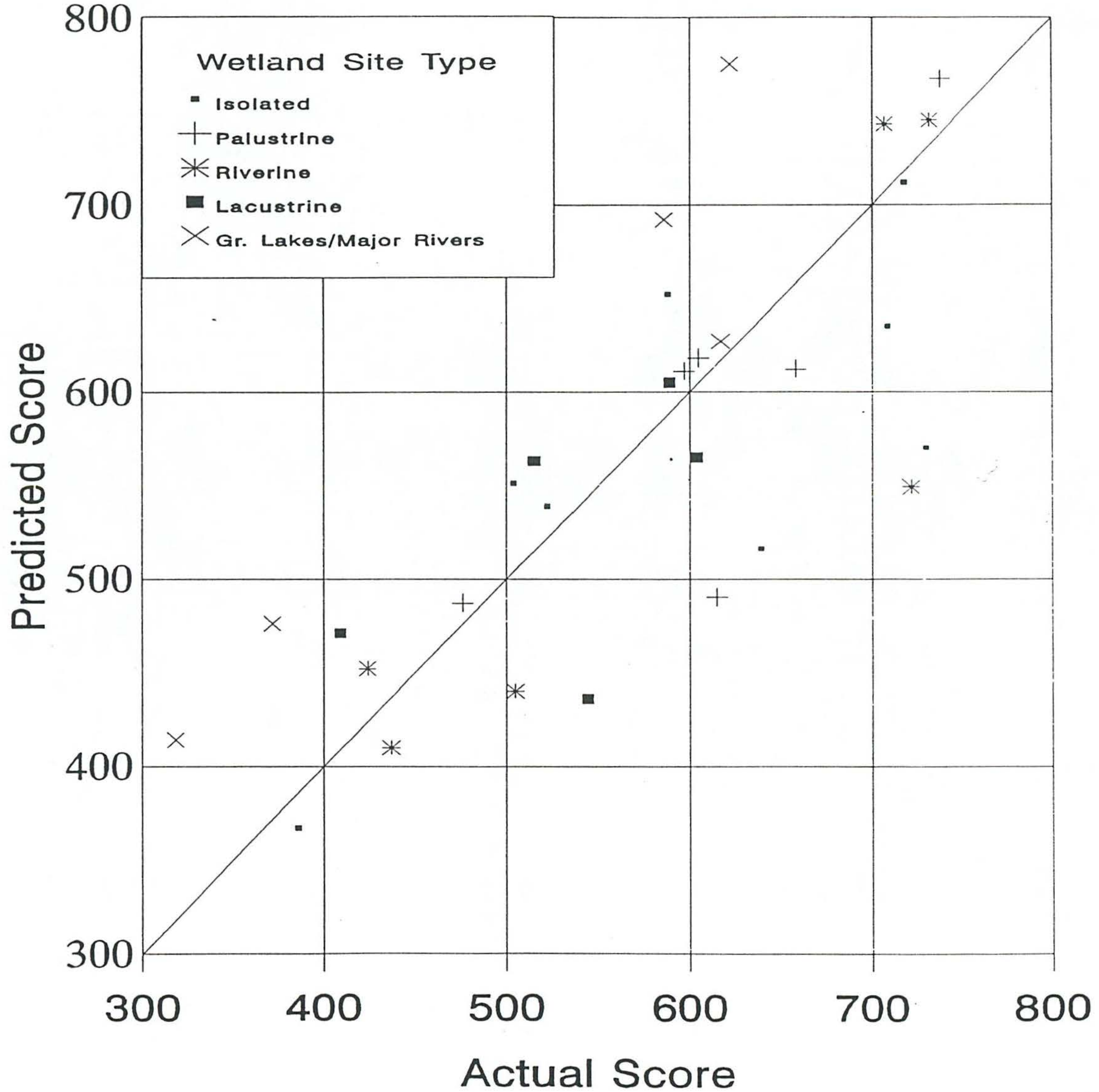
$$2.639(32.850) + 60(1.424) + 68(2.002) + 70(1.531) + 8(3.596) + 26(2.581) + 8(6.494) + 23.297 = 587$$

In linear multiple regression, the adjusted R^2 indicates the strength of the relationship between xi and yi in the data. This is referred to as the fitting accuracy of the model. Ultimately however, the usefulness of a model depends on how accurately it can predict the score of a wetland that is not included in the sample population. This is known as the model's predictive accuracy.

As additional wetland evaluations are not currently available, the predictive accuracy of the models presented can be tested by removing a number of wetlands from the sample, recalculating the regression equation and observing how closely the revised model predicts the score of the removed wetlands. This procedure was applied to model S1D. Thirty cases were removed which varied in size, site type and total score. This reduced the size of N to 93. After recalculating the equation, the value of the adjusted R^2 increased to 0.825 from its original value of 0.770.

The results of this test are plotted in Figure 1. It appears that the model is reasonably accurate with regards to this sample and is especially strong at predicting the score of palustrine wetlands. The plot also reveals that the model may underpredict the score of an individual wetland by as much as 20 percent. This is evident from the data points which are below and furthest away from the line. In 60 percent of the cases however, the model overestimated the wetland scores. Such a result is desirable as it reduces the likelihood that a provincially significant wetland will be mistakenly undervalued and eliminated from further study.

FIGURE 1: PREDICTIVE ACCURACY OF MODEL S1D
USING 30 TEST WETLANDS
ADJUSTED R2 = 0.825, N=93



Part B: Northern Ontario Wetland Analysis

Variables considered for inclusion in the northern models were selected from the Ontario Wetland Evaluation System Northern Manual (OMNR, 1993a). Generally, the attributes contained in the northern and southern manuals are the same with the exception of the Hydrological component; for example, the attributes which score the water quality improvement function are entirely different. Consequently, the northern Group "A" Hydrological variables are listed separately in Table 4. The northern variables of the other three components, as well as their shorthand name and classification, are the same as the southern variables listed in Table 1.

Table 4: Northern Group "A" Hydrological Variables

VARIABLE	FULL NAME	SECTION IN NORTHERN MANUAL
FLOOD	Flood Attenuation	3.1
WST	Wetland Site Type	3.2.1
TDWQI	Downstream Water Quality Improvement (Total)	3.3
SEC	Shoreline Erosion Control	3.5

The statistics for the northern variables are located in the second part of Appendix A. As the output indicates, the only variable that was significantly skewed was SIZE1. This condition was rectified by performing a log transformation which produced the variable SZ1LOG. The statistics also report that the mean size of these wetlands is approximately 310 hectares with a mean score of 603 points. Both of these values are somewhat larger than those obtained from the southern data.

As the northern data consist of 66 randomly evaluated wetlands, up to five independent variables were included in the models in order to display a broad range of adjusted R^2 values and to illustrate the amount this statistic decreases in subsequent models. Initially model S1D, the preferred five variable model developed for the southern data, was applied to the northern data set. The resulting adjusted R^2 value dropped from 0.770 to 0.672. Comparatively lower values were also reported when models S1E, S1F and S1G were applied. By testing different combinations of predictor variables, a higher adjusted R^2 value of 0.691 was obtained by replacing the variables REC and FLOOD with AESTH and OWNER. These are the terms of model N1A which is listed in Table 5. Overall the N1 models, which are listed in the top section of Table 5, best fit the northern data.

Table 5: Northern Models (N=66)

N1 Models							
Model	Constant	RTOT1	SZ1LOG	AESTH	OPWAT	OWNER	Adj R^2
A	350.387	1.9393	38.248	-12.879	3.062	7.475	0.691
B	385.755	2.006	37.109	-10.716	3.073	-	0.671
C	436.669	2.275	37.469	-12.729	-	-	0.627
D	346.327	2.010	39.618	-	-	-	0.571

N2 Models							
Model	Constant	SZ1LOG	NOTYPE	OPWAT	EDTOT	FLOOD	Adj R^2
A	215.172	34.888	4.163	5.142	3.257	0.704	0.694
B	267.043	32.554	5.069	3.990	2.616	-	0.652
C	286.579	34.915	4.535	3.877	-	-	0.616
D	322.230	35.702	5.072	-	-	-	0.541

An alternative series of "next best" northern models was developed to facilitate the prediction of wetland scores. This series, labelled the N2 models, is located in the lower half of Table 5. The complete statistical data for all of the northern models are listed in Appendix B. Although the five-variable model of the N1 series has a slightly lower adjusted R^2 value (0.691) than its N2 counterpart (0.694), the values of the subsequent N1 models exceed those of N2 in each case. The N2 models however are considered to be reliable since the relative difference in the adjusted R^2 values between the corresponding models varies by no more than 5.3 percent, which is the variation between models N2D and N1D (0.541 versus 0.571).

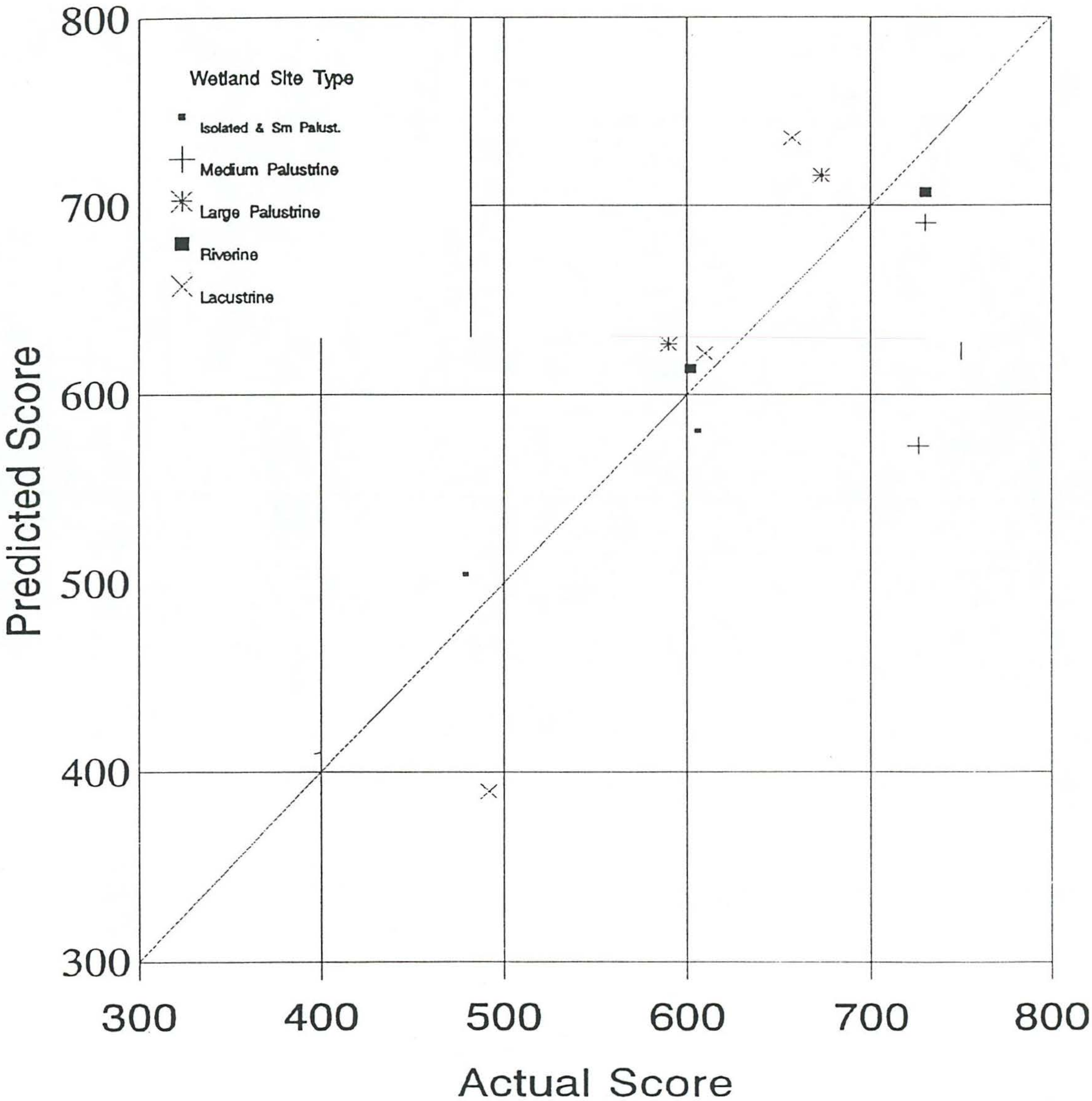
The correlation matrix in Appendix C indicates that multicollinearity is not a problem with this data given the low values between the variables. Furthermore, diagnostic tests conducted on model N1A indicate that assumptions of regression analysis regarding the distribution of the errors have not been violated. These tests are shown in Appendix D.

Test of Model Accuracy

In order to assess the predictive accuracy of model N1A, it underwent the same test that was performed on model S1D. In this case, the regression equation was recalculated after 12 wetlands that ranged in size, site type and total score were removed from the sample. In the new equation, the adjusted R^2 value increased to 0.730. The results of this test are plotted in Figure 2. As the northern data set is relatively small, the results of this test must be interpreted cautiously. They do however provide a general overview of the model's performance. Based on these test cases, the model appears to be reasonably accurate, especially in its ability to predict the score of riverine wetlands.

Unlike the southern model (S1D) which overestimated the score of 60 percent of the wetlands tested, the proportion of wetlands underestimated by this model was 50 percent. Such a result is not unexpected however, given the random distribution of the prediction error. The plot also indicates that the model may underestimate the score of an individual wetland by as much as 20 percent.

FIGURE 2: PREDICTIVE ACCURACY OF MODEL N1A
USING 12 TEST WETLANDS
ADJUSTING R2 = 0.730, N=54



Discussion and Conclusions

The models developed in this paper indicate that there are strong relationships between the total wetland score and key independent variables. This is demonstrated by the high values of adjusted R^2, together with the substantial and statistically significant values of the regression coefficients. The P-value, which is 0 in all cases, indicates that each model is significant at the 0.01 critical level. Furthermore, the P(2 TAIL) statistic confirms that in every model, all of the independent variables are significant at the 0.05 critical level. In fact, the majority are significant even at the 0.01 level.

Although the northern and southern models are both statistically valid, there are differences in their adjusted R^2 coefficients. The preferred five-variable northern model (model N1A) could account for nearly 70 percent of the variability in the data and will have useful applications. This result however, is slightly poorer than the adjusted R^2 value obtained for the preferred five-variable southern model (model S1D) which could account for 77 percent of the data variability. As additional data becomes available, it is possible that the northern models could be further improved. Confidence in model N1A however is strengthened by the fact that it shares three out of five predictor variables (SZ1LOG, RTOT1 and OPWAT) with model S1D, which was developed from a data set that contained nearly twice as many observations.

The R^2 values of model S1D and model N1A exceed the R^2 value of the preferred model developed in a similar study by Scheifele and Mulamoottil (1987). In that study, regression models were developed from 437 southern Ontario wetlands that were evaluated under the 1984 evaluation system (Environment Canada and MNR, 1984). That model had an R^2 value of 0.609 and included the following independent variables: number of vegetation communities, log of the wetland size, proximity to urban centres, number of wetland types and percentage of organic soil. The Scheifele & Mulamoottil study differed from this one in that it utilized raw evaluation data rather than data converted into wetland scores.

Although the estimates provided by predictive models serve as a useful guideline, they do not replace the need to undertake full, on-site evaluations. Both the northern and southern evaluation manuals include many important values and functions that are not considered by these models. Furthermore, as evident from Figures 1 and 2, the predicted score may underestimate the actual score of an individual wetland by as much as 20 percent. Although this amount is not extensive, it highlights the fact that caution and professional judgement must be applied in the interpretation of the results.

Secondly, these models are limited by the fact that they will not identify wetlands that are provincially significant because they score 200 or more points on either the Biological or Special Features Components. Thirdly, in order for predictive models to make estimates that are meaningful, the individuals using them must be skilled at delineating wetland boundaries and identifying features such as wetland type, site type and predominant vegetation forms from maps and aerial photographs. Finally, since the predicted scores merely indicate a wetland's potential value, they cannot be used as proof of provincial significance for tax rebate purposes or as evidence at Ontario Municipal Board hearings.

Within these constraints however, the proper use of regression models can serve as a useful planning tool that can improve the quality of land use decisions. For example, by providing a means of prestratifying wetlands based on their potential value, these models will enable limited funds to be directed to those wetlands that are the most likely to be provincially significant. In addition, predictive models can provide important information to other resource management applications such as the evaluation of utility and transportation corridor options. They will also assist in the identification of values for Forest Management Plans in cases where wetlands are involved.

The models produced by this research are a function of the data which is currently available. As subsequent wetlands are evaluated, it is expected that they will need to be reevaluated and modified. In addition, the evaluation system upon which they are based will be revised as our knowledge of wetland science continues to develop. Consequently, similar studies need to be undertaken in the future in order to ensure that these models remain as accurate and up-to-date as possible. This is important given that the information that they provide can play a significant role in the identification and protection of this valuable resource.

Adjusted Squared Multiple R - is the Squared Multiple R "adjusted" for the number of independent variables, or degrees of freedom, used in the model.

Coefficient - indicates the degree to which an independent variable influences the dependent variable.

F - Ratio - tests the "goodness of fit" of the entire model. It equals the Mean Square of Regression divided by the Mean Square of Residual.

P - the probablity of getting results at least as extreme as this in the data assuming that the null hypothesis is true. As the computed value P was 0 in all cases, every model presented in this study is significant at the 0.01 critical level.

P(2 Tail) - is the probability of obtaining, under the null hypothesis of no association, a t-value as extreme as the one observed. When P is less than 0.05, the null hypothesis is rejected at the five percent critical level.

Residual Error - measures the unexplained or random variability in the data. It is equal to the square root of the residual mean square.

Squared Multiple R - measures the percent of variation in the dependent variable which is explained by variations in the independent variables taken together. In each model, TOTAL (total wetland score) is the dependent variable. For example, in model S1A, 81.7 percent of the variation in TOTAL can be explained by the combined effect of the other variables. The remaining variation is attributed to the residual error.

Standardized Coefficient (STD COEFF) - is the coefficient measured in standard deviations.

Tolerance - measures the degree of independence among predictor variables. Its value is 1 minus the squared multiple correlation between a predictor and other predictors included in the model. A high value (i.e. close to 1) indicates that intercorrelation is not a problem.

Cohen J. and P. Cohen. 1975. Applied Regression Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Hillsdale New Jersey.

Environment Canada and Ontario Ministry of Natural Resources. 1984. An Evaluation System for Wetlands of Ontario South of the Precambrian Shield. 2nd ed.

Draper N.R. and H. Smith. 1981. Applied Regression Analysis. John Wiley and Sons. New York.

Ontario Ministry of Natural Resources. 1993a. Ontario Wetland Evaluation System Northern Manual. NEST Technical Manual TM-001.

Ontario Ministry of Natural Resources. 1993b. Ontario Wetland Evaluation System Southern Manual. NEST Technical Manual TM-002.

Ontario Ministry of Natural Resources and Ontario Ministry of Municipal Affairs. 1992. Wetlands: A Statement of Ontario Government Policy Issued Under Authority of Section 3 of the Planning Act 1983. Toronto.

Rawlings, J.O. 1988. Applied Regression Analysis: A Research Tool. Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.

Scheifele, G.W. and G. Mulamoottil 1987. Predictive Models Applicable to Ontario's Wetland Evaluation System. Proceedings of Symposium '87 Wetlands/Peatlands, Edmonton Alberta. pp. 267-273.

Snell, E.A. 1987. Wetlands Distribution and Conversion in Southern Ontario. Working paper No. 48. Inland Waters and Lands Directorate, Environment Canada. Ottawa.

Weisberg, S. 1985. Applied Linear Regression. John Wiley & Son. New York.

Wilkinson L. 1989. SYSTAT: The System for Statistics, SYSTAT Inc., Evanston Illinois.

Appendix A: Group "A" Variable Statistics

1. Southern Ontario Data

Total Observations for Each Variable = 123

	SIZE1	WLTYPE	SITE	NOTYPE	HABDIV
MINIMUM	1.200	5.000	1.000	9.000	2.000
MAXIMUM	2354.400	15.000	4.000	30.000	36.000
MEAN	246.638	11.699	2.366	4.122	6.935
STANDARD DEV	375.893	2.028	0.926	4.375	2.757
SKEWNESS(G1)	3.126	-0.415	0.584	1.730	9.541
MEDIAN	105.000	11.000	2.000	13.000	7.000

	WPROX	OPWAT	EVP	REC	AESTH
MINIMUM	0.000	0.000	4.000	0.000	0.000
MAXIMUM	8.000	30.000	50.000	80.000	10.000
MEAN	6.285	11.285	29.951	32.260	6.041
STANDARD DEV	2.578	7.478	10.315	24.480	1.870
SKEWNESS(G1)	-1.337	0.631	-0.573	0.689	-0.361
MEDIAN	8.000	8.000	31.000	28.000	7.000

	EDTOT	HPROX	OWNER	FLOOD	SEC
MINIMUM	0.000	10.000	4.000	0.000	0.000
MAXIMUM	36.000	40.000	8.000	100.000	15.000
MEAN	7.447	17.935	5.049	71.081	6.260
STANDARD DEV	9.853	8.049	1.470	39.588	6.097
SKEWNESS(G1)	1.484	1.529	1.050	-0.984	0.229
MEDIAN	5.000	16.000	4.000	97.000	7.000

	RTOT1	AGE	TOTAL
MINIMUM	0.000	0.000	318.000
MAXIMUM	160.000	25.000	853.000
MEAN	71.432	3.285	577.707
STANDARD DEV	42.366	3.769	119.691
SKEWNESS(G1)	0.705	3.781	0.066
MEDIAN	60.000	3.000	586.000

Statistics for Log Transformed Variables

	SZ1LOG	HABLOG	AGLOG
MINIMUM	0.182	0.693	-2.996
MAXIMUM	7.764	3.584	3.221
MEAN	4.614	1.905	0.552
STANDARD DEV	1.478	0.221	1.535
SKEWNESS(G1)	-0.450	1.747	-1.484
MEDIAN	4.654	1.946	1.115

2. Northern Ontario Data

Total Observations for Each Variable = 66

	SIZE1	WLTYPE	SITE	NOTYPE	HABDIV
MINIMUM	1.000	7.000	1.000	9.000	2.000
MAXIMUM	4125.000	15.000	7.000	30.000	6.000
MEAN	310.045	11.409	2.621	19.682	5.455
STANDARD DEV	603.491	2.127	1.134	7.160	0.948
SKEWNESS(G1)	4.614	-0.352	1.421	0.378	-1.782
MEDIAN	143.500	12.000	2.000	20.000	6.000

	WPROX	OPWAT	EVP	REC	AESTH
MINIMUM	0.000	0.000	0.000	0.000	3.000
MAXIMUM	8.000	30.000	51.000	80.000	11.000
MEAN	6.970	15.470	29.030	31.894	6.985
STANDARD DEV	2.075	7.736	9.853	23.035	2.072
SKEWNESS(G1)	-1.864	0.726	-0.575	0.651	0.062
MEDIAN	8.000	14.000	30.000	30.000	7.000

	EDTOT	HPROX	OWNER	FLOOD	WST
MINIMUM	0.000	0.000	4.000	0.000	0.000
MAXIMUM	33.000	40.000	17.000	100.000	20.000
MEAN	7.106	17.500	6.364	51.394	11.985
STANDARD DEV	8.170	10.796	2.264	35.516	7.922
SKEWNESS(G1)	1.546	1.215	1.569	-0.278	-0.270
MEDIAN	5.000	12.000	6.000	65.500	14.500

	TDWQI	SEC	RTOT1	AGE	TOTAL
MINIMUM	21.000	0.000	0.000	0.000	324.000
MAXIMUM	100.000	15.000	70.000	20.000	799.000
MEAN	54.318	6.227	33.333	5.636	602.939
STANDARD DEV	13.284	5.262	9.001	5.445	105.721
SKEWNESS(G1)	0.399	0.232	0.070	1.140	-0.494
MEDIAN	54.000	8.000	35.000	3.000	603.000

Statistics for Variable SIZE1 After Log Transformation

	SZ1LOG
MINIMUM	0.000
MAXIMUM	8.325
MEAN	4.786
STANDARD DEV	1.499
SKEWNESS(G1)	-0.560
MEDIAN	4.966

For an explanation of the statistical terms included in the following data, refer to the glossary.

1. SOUTHERN ONTARIO MODELS

MODEL S1A

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.911 SQUARED MULTIPLE R: 0.829 ADJUSTED SQUARED MULTIPLE R: .817 RESIDUAL ERROR: 51.196

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	35.462	31.253	0.000	.	1.135	0.259
SZ1LOG	33.442	3.631	0.413	0.7463042	9.211	0.000
WPROX	6.316	2.021	0.136	0.7913076	3.125	0.002
OPWAT	3.582	0.685	0.224	0.8197636	5.232	0.000
REC	1.693	0.259	0.346	0.5346358	6.539	0.000
EDTOT	1.751	0.573	0.144	0.6741459	3.056	0.003
HPROX	2.141	0.629	0.144	0.8376390	3.403	0.001
FLOOD	1.533	0.157	0.507	0.5533261	9.737	0.000
RTOT1	1.299	0.122	0.460	0.8009827	10.630	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1448962.606	8	181120.326	69.102	0.000
RESIDUAL	298798.858	114	2621.043		

MODEL S1B

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.903 SQUARED MULTIPLE R: 0.815 ADJUSTED SQUARED MULTIPLE R: .804 RESIDUAL ERROR: 53.020

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	23.297	32.103	0.000	.	0.726	0.469
SZ1LOG	32.850	3.755	0.406	0.7484382	8.749	0.000
WPROX	6.494	2.092	0.140	0.7919645	3.104	0.002
OPWAT	3.596	0.709	0.225	0.8198026	5.073	0.000
REC	2.002	0.247	0.409	0.6305928	8.108	0.000
HPROX	2.581	0.634	0.174	0.8838991	4.069	0.000
FLOOD	1.531	0.163	0.506	0.5533325	9.392	0.000
RTOT1	1.424	0.119	0.504	0.9018635	11.939	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1424480.791	7	203497.256	72.390	0.000
RESIDUAL	323280.673	115	2811.136		

MODEL S1C

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.894 SQUARED MULTIPLE R: 0.800
ADJUSTED SQUARED MULTIPLE R: .789 RESIDUAL ERROR: 54.958

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	77.151	27.999	0.000	.	2.756	0.007
SZ1LOG	33.112	3.891	0.409	0.7488174	8.510	0.000
OPWAT	3.476	0.734	0.217	0.8222770	4.737	0.000
REC	2.094	0.254	0.428	0.6398651	8.242	0.000
HPROX	2.191	0.644	0.147	0.9200293	3.400	0.001
FLOOD	1.370	0.160	0.453	0.6159482	8.553	0.000
RTOT1	1.461	0.123	0.517	0.9106456	11.869	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1397395.597	6	232899.266	77.109	0.000
RESIDUAL	350365.866	116	3020.395		

MODEL S1D

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.883 SQUARED MULTIPLE R: 0.780 ADJUSTED SQUARED MULTIPLE R: .770 RESIDUAL ERROR: 57.384

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	131.674	23.964	0.000	.	5.495	0.000
SZ1LOG	30.628	3.990	0.378	0.7761739	7.676	0.000
OPWAT	3.045	0.755	0.190	0.8475186	4.035	0.000
REC	2.181	0.264	0.446	0.6464460	8.264	0.000
FLOOD	1.333	0.167	0.441	0.6187063	7.992	0.000
RTOT1	1.473	0.128	0.521	0.9114082	11.467	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1362483.581	5	272496.716	82.751	0.000
RESIDUAL	385277.88	117	3292.973		

MODEL S1E

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.865 SQUARED MULTIPLE R: 0.749 ADJUSTED SQUARED MULTIPLE R: .740 RESIDUAL ERROR: 60.987

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	168.414	23.559	0.000	.	7.149	0.000
SZ1LOG	32.578	4.210	0.402	0.7877204	7.739	0.000
REC	2.194	0.281	0.449	0.6465362	7.822	0.000
FLOOD	1.113	0.168	0.368	0.6928377	6.643	0.000
RTOT1	1.527	0.136	0.541	0.9215037	11.248	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1308864.488	4	327216.122	87.974	0.000
RESIDUAL	438896.975	118	3719.466		

MODEL S1F

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.809 SQUARED MULTIPLE R: 0.655 ADJUSTED SQUARED MULTIPLE R: .646 RESIDUAL ERROR: 71.187

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	224.745	25.656	0.000	.	8.760	0.000
SZ1LOG	41.130	4.678	0.508	0.8689837	8.792	0.000
REC	1.222	0.279	0.250	0.8884191	4.374	0.000
RTOT1	1.733	0.154	0.613	0.972126	11.232	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1144725.223	3	381575.074	75.298	0.000
RESIDUAL	603036.241	119	5067.531		

MODEL S1G

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.774 SQUARED MULTIPLE R: 0.599 ADJUSTED SQUARED MULTIPLE R: .593 RESIDUAL ERROR: 76.376

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	234.384	27.424	0.000	.	8.547	0.000
SZ1LOG	47.808	4.745	0.590	0.9725800	10.077	0.000
RTOT1	1.718	0.165	0.608	0.9725800	10.383	0.000

ANALYSIS OF VARIANCE

RESOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1047772.220	2	523886.110	89.810	0.000
RESIDUAL	699989.243	120	5833.244		

MODEL S2A

DEP VAR: T OTAL N: 123 MULTIPLE R: 0.910 SQUARED MULTIPLE R: 0.827 ADJUSTED SQUARED MULTIPLE R: .815 RESIDUAL ERROR: 51.436

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	78.867	26.628	0.000	.	2.962	0.004
EVP	2.140	0.729	0.184	0.3832548	2.935	0.004
RTOT1	1.346	0.122	0.476	0.8051817	10.986	0.000
REC	1.748	0.259	0.357	0.5393576	6.748	0.000
FLOOD	1.356	0.150	0.448	0.6145316	9.034	0.000
OPWAT	3.249	0.691	0.203	0.8129795	4.704	0.000
SZ1LOG	22.468	5.290	0.277	0.3548651	4.247	0.000
EDTOT	1.744	0.576	0.144	0.6738889	3.029	0.003
HPROX	1.815	0.620	0.122	0.8712005	2.928	0.004

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1446151.840	8	180768.980	68.326	0.000
RESIDUAL	301609.623	114	2645.698		

MODEL S2B

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.902 SQUARED MULTIPLE R: 0.814 ADJUSTED SQUARED MULTIPLE R: .803 RESIDUAL ERROR: 53.103

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	124.301	22.339	0.000	.	5.564	0.000
EVP	2.064	0.752	0.178	0.3837519	2.743	0.007
RTOT1	1.327	0.126	0.470	0.8073546	10.509	0.000
REC	1.749	0.267	0.358	0.5393588	6.541	0.000
FLOOD	1.329	0.155	0.440	0.6167682	8.596	0.000
OPWAT	2.917	0.703	0.182	0.8355530	4.147	0.000
SZ1LOG	21.052	5.438	0.260	0.3578556	3.871	0.000
EDTOT	2.129	0.579	0.175	0.7110678	3.680	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1423475.123	7	203353.589	72.114	0.000
RESIDUAL	324286.340	115	2819.881		

MODEL S2C

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.890 SQUARED MULTIPLE R: 0.793 ADJUSTED SQUARED MULTIPLE R: .782 RESIDUAL ERROR: 55.899

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	124.020	23.515	0.000	.	5.274	0.000
EVP	2.139	0.792	0.184	0.3840381	2.702	0.008
RTOT1	1.484	0.125	0.525	0.9104951	11.851	0.000
REC	2.147	0.257	0.439	0.6448663	8.340	0.000
FLOOD	1.312	0.163	0.434	0.6173043	8.066	0.000
OPWAT	2.820	0.740	0.176	0.8367281	3.811	0.000
SZ1LOG	19.350	5.704	0.239	0.3604627	3.392	0.001

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1385293.050	6	230882.175	73.889	0.000
RESIDUAL	362468.413	116	3124.728		

MODEL S2D

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.879 SQUARED MULTIPLE R: 0.772 ADJUSTED SQUARED MULTIPLE R: .762 RESIDUAL ERROR: 58.356

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	143.866	23.777	0.000	.	6.051	0.000
EVP	4.105	0.563	0.354	0.8269383	7.288	0.000
RTOT1	1.430	0.130	0.506	0.9250743	11.031	0.000
REC	2.365	0.260	0.484	0.6877602	9.089	0.000
FLOOD	1.416	0.167	0.468	0.6397680	8.486	0.000
OPWAT	2.818	0.772	0.176	0.8367283	3.649	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1349332.307	5	269866.461	79.247	0.000
RESIDUAL	398429.156	117	3405.377		

MODEL S2E

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.864 SQUARED MULTIPLE R: 0.746 ADJUSTED SQUARED MULTIPLE R: .737 RESIDUAL ERROR: 61.325

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	176.203	23.186	0.000	.	7.600	0.000
EVP	4.444	0.584	0.383	0.8500626	7.612	0.000
RTOT1	1.482	0.135	0.524	0.9360844	10.938	0.000
REC	2.371	0.273	0.485	0.6877865	8.670	0.000
FLOOD	1.211	0.165	0.401	0.7212322	7.335	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1303995.467	4	325998.867	86.685	0.000
RESIDUAL	443765.996	118	3760.729		

MODEL S2F

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.794 SQUARED MULTIPLE R: 0.630 ADJUSTED SQUARED MULTIPLE R: .621 RESIDUAL ERROR: 73.684

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	249.201	25.161	0.000	.	9.904	0.000
EVP	5.460	0.681	0.471	0.9008383	8.013	0.000
RTOT1	1.689	0.159	0.598	0.9788689	10.614	0.000
REC	1.373	0.285	0.281	0.9138743	4.818	0.000

ANALYSIS OF VARIANCE SOURCE

	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	1101678.679	3	367226.226	67.638	0.000
RESIDUAL	646082.784	119	5429.267		

MODEL S2G

DEP VAR: TOTAL N: 123 MULTIPLE R: 0.747 SQUARED MULTIPLE R: 0.558 ADJUSTED SQUARED MULTIPLE R: .551 RESIDUAL ERROR: 80.214

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	267.515	27.077	0.000	.	9.880	0.00
EVP	6.394	0.711	0.551	0.9801488	8.991	0.000
RTOT1	1.662	0.173	0.588	0.9801488	9.597	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	975650.529	2	487825.264	75.817	0.000
RESIDUAL	772110.935	120	6434.258		

2. Northern Ontario Models

MODEL N1A

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.845 SQUARED MULTIPLE R: 0.715 ADJUSTED SQUARED MULTIPLE R: .691 RESIDUAL ERROR: 58.788

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	350.387	42.257	0.000	.	8.292	0.000
RTOT1	1.939	0.425	0.348	0.8162584	4.564	0.000
SZ1LOG	38.248	5.235	0.542	0.8632828	7.306	0.000
AESTH	-12.897	3.772	-0.253	0.8707953	-3.419	0.001
OPWAT	3.062	0.977	0.224	0.9300066	3.133	0.003
OWNER	7.475	3.377	0.160	0.9092768	2.213	0.031

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	519135.270	5	103827.054	30.042	0.000
RESIDUAL	207362.487	60	3456.041		

MODEL N1B

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.831 SQUARED MULTIPLE R: 0.691 ADJUSTED SQUARED MULTIPLE R: .671 RESIDUAL ERROR: 60.638

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	385.755	40.350	0.000	.	9.560	0.000
RTOT1	2.006	0.437	0.361	0.8204672	4.591	0.000
SZ1LOG	37.109	5.374	0.526	0.8717075	6.906	0.000
AESTH	-10.716	3.755	-0.210	0.9345551	-2.854	0.006
OPWAT	3.073	1.008	0.225	0.9300322	3.049	0.003

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	502203.666	4	125550.916	34.145	0.000
RESIDUAL	224294.092	61	3676.952		

MODEL N1C

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.803 SQUARED MULTIPLE R: 0.644 ADJUSTED SQUARED MULTIPLE R: .627 RESIDUAL ERROR: 64.567

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	436.669	39.112	0.000	.	11.165	0.000
RTOT1	2.275	0.456	0.409	0.8553300	4.992	0.000
SZ1LOG	37.469	5.720	0.531	0.8721298	6.550	0.000
AESTH	-12.729	3.936	-0.249	0.9643673	-3.234	0.002

ANALYSIS OF VARIANCE SOURCE

	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	468026.391	3	156008.797	37.422	0.000
RESIDUAL	258471.367	62	4168.893		

MODEL N1D

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.764 SQUARED MULTIPLE R: 0.584 ADJUSTED SQUARED MULTIPLE R: .571 RESIDUAL ERROR: 69.244

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	346.327	29.356	0.000	.	11.798	0.000
RTOT1	2.010	0.481	0.361	0.8840561	4.180	0.000
SZ1LOG	39.618	6.093	0.562	0.8840561	6.502	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	424430.485	2	212215.242	44.260	0.000
RESIDUAL	302067.273	63	4794.719		

MODEL N2A

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.847 SQUARED MULTIPLE R: 0.717 ADJUSTED SQUARED MULTIPLE R: .694 RESIDUAL ERROR: 58.500

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	215.172	33.643	0.000	.	6.396	0.000
SZ1LOG	34.888	5.777	0.495	0.7019136	6.039	0.000
NOTYPE	4.163	1.243	0.282	0.6643634	3.348	0.001
OPWAT	5.142	1.022	0.376	0.8427417	5.033	0.000
EDTOT	3.257	0.926	0.252	0.9194030	3.516	0.001
FLOOD	0.704	0.230	0.236	0.7892492	3.060	0.003

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	521165.725	5	104233.145	30.458	0.000
RESIDUAL	205332.033	60	3422.201		

MODEL N2B

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.821 SQUARED MULTIPLE R: 0.673 ADJUSTED SQUARED MULTIPLE R: .652 RESIDUAL ERROR: 62.380

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	267.043	30.987	0.000	.	8.618	0.000
SZ1LOG	32.554	6.106	0.462	0.7143657	5.331	0.000
NOTYPE	5.069	1.288	0.343	0.7043398	3.937	0.000
OPWAT	3.990	1.013	0.292	0.9751753	3.939	0.000
EDTOT	2.616	0.962	0.202	0.9689850	2.719	0.009

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	489127.803	4	122281.951	31.424	0.000
RESIDUAL	237369.954	61	3891.311		

MODEL N2C

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.796 SQUARED MULTIPLE R: 0.634 ADJUSTED SQUARED MULTIPLE R: .616 RESIDUAL ERROR: .. 65.518

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	286.579	31.658	0.000	.	9.052	0.000
SZ1LOG	34.915	6.348	0.495	0.7291041	5.500	0.000
NOTYPE	4.535	1.337	0.307	0.7211036	3.393	0.001
OPWAT	3.877	1.063	0.284	0.9768036	3.648	0.001

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSIO	460358.645	3	153452.882	35.749	0.000
RESIDUAL	266139.112	62	4292.566		

MODEL N2D

DEP VAR: TOTAL N: 66 MULTIPLE R: 0.745 SQUARED MULTIPLE R: 0.555 ADJUSTED SQUARED MULTIPLE R: .541 RESIDUAL ERROR: 71.633

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	332.230	31.795	0.000	.	10.449	0.000
SZ1LOG	35.702	6.937	0.506	0.7299472	5.147	0.000
NOTYPE	5.072	1.452	0.343	0.7299472	3.492	0.001

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	403228.076	2	201614.038	39.291	0.000
RESIDUAL	323269.682	63	5131.265		

Appendix C: Matrix of Spearman Correlation Coefficients

1. Southern Ontario Data

Number of Observations = 123

	SZ1LOG	WPROX	OPWAT	EVP	REC
SZ1LOG	1.000				
WPROX	0.021	1.000			
OPWAT	0.077	0.149	1.000		
EVP	0.813	-0.001	0.090	1.000	
REC	0.336	0.355	0.234	0.272	1.000
EDTOT	0.115	0.208	0.098	0.122	0.405
HPROX	-0.165	-0.026	-0.128	-0.168	0.122
FLOOD	0.098	-0.445	-0.294	0.084	-0.486
RTOT1	-0.094	0.000	0.021	-0.124	-0.047
TOTAL	0.525	0.078	0.17	0.482	0.416

	EDTOT	HPROX	FLOOD	RTOT1	TOTAL
EDTOT	1.000				
HPROX	0.148	1.000			
FLOOD	-0.160	-0.216	1.000		
RTOT1	0.342	0.081	0.145	1.000	
TOTAL	0.482	-0.010	0.162	0.511	1.000

2. Northern Ontario Data

Number of Observations = 66

	SZ1LOG	NOTYPE	OPWAT	EDTOT	FLOOD
SZ1LOG	1.000				
NOTYPE	0.526	1.000			
OPWAT	0.026	0.159	1.000		
EDTOT	-0.014	-0.047	0.030	1.000	
FLOOD	0.033	0.126	-0.371	-0.272	1.000
RTOT1	0.376	0.771	0.201	-0.072	0.244
AESTH	-0.016	0.106	-0.201	-0.226	0.182
OWNER	-0.096	0.025	0.027	0.226	-0.087
TOTAL	0.609	0.620	0.345	0.152	0.094

	RTOT1	AESTH	OWNER	TOTAL
RTOT1	1.000			
AESTH	0.152	1.000		
OWNER	0.123	0.213	1.000	
TOTAL	0.544	-0.255	-0.019	1.000

In order to ensure that the residual are from a normal distribution, diagnostic plots were developed for the southern and northern data using models S1A and N1A as test cases for the S1 and N1 model series. Figures 3 to 6 apply to the southern data while Figures 7 to 10 pertain to the north.

Figure 3 is a probability plot using the southern data, where total wetland score (TOTAL) is plotted against the expected values of a normal distribution. If the data is normally distributed, the plotted values will lie on a straight diagonal line as is the case with this figure. Similarly, the residual errors, as plotted in Figure 4, should also be normally distributed and therefore follow the same pattern.

Figure 5 is a plot of the studentized residuals (STUDENT) against the estimated values (ESTIMATE). Studentized residuals follow approximately a t distribution, therefore large values (greater than 3) could be possible problems. As the figure indicates, all values lie within -3 and +3.

Finally the model was tested to ensure that all members of the sampled population are described by the same linear model. The Cook's distance (COOK) in Figure 6 measures the degree of influence each sample estimate has on the coefficient estimate. A Cook's distance greater than 0.8 to 1.0 is considered large (Wilkinson, 1989). Observations that vary significantly from the average of all the independent variables or have large residuals themselves will have a large Cook's distance. This is not a problem with the data set since the largest Cook's value is about 0.12. The northern plots (yield similar results (Figures 7 to 10) although in Figure 10 the maximum Cook's distance is just under 0.6. This result however is still well under the critical value of 0.8 to 1.0. As the models tested meet the criteria of the test plots, the assumptions which are based upon them are strengthened.

Diagnostic Plots for Model N1A

Figure 3

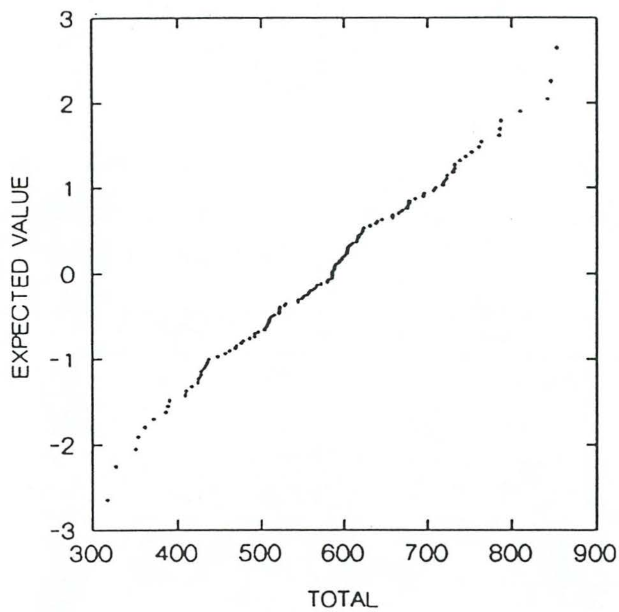


Figure 4

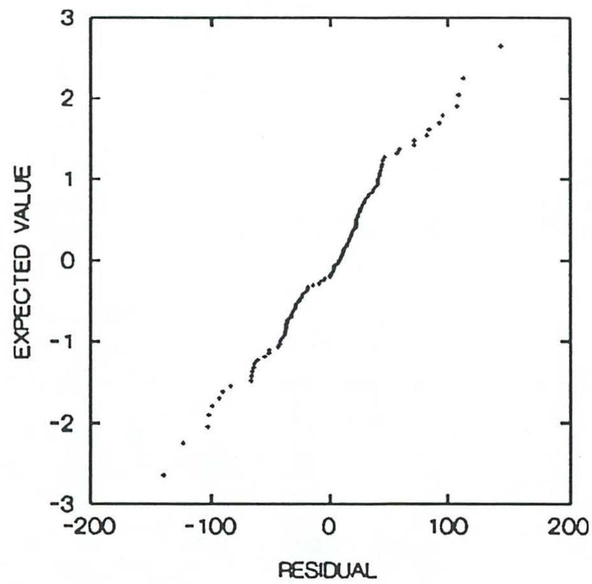


Figure 5

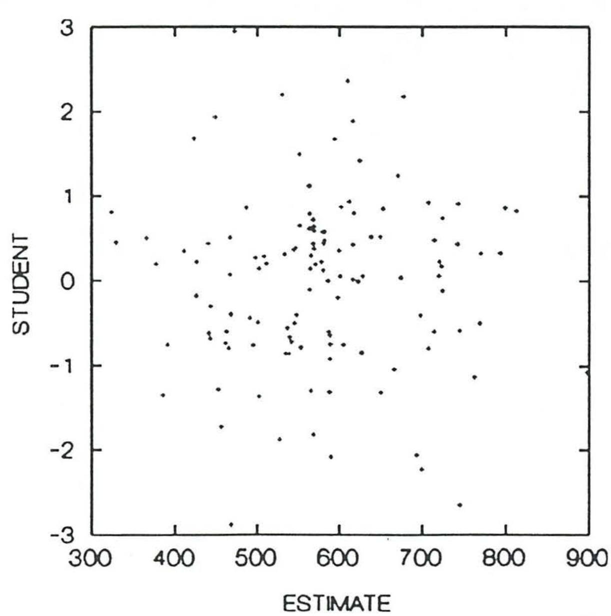


Figure 6

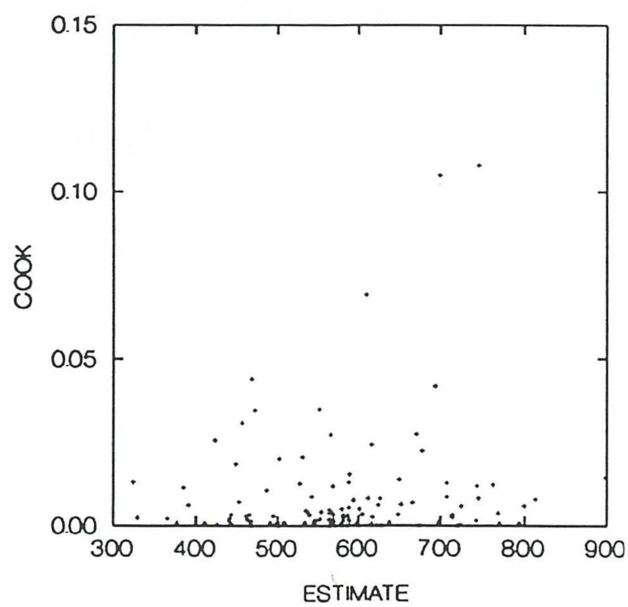


FIGURE 7

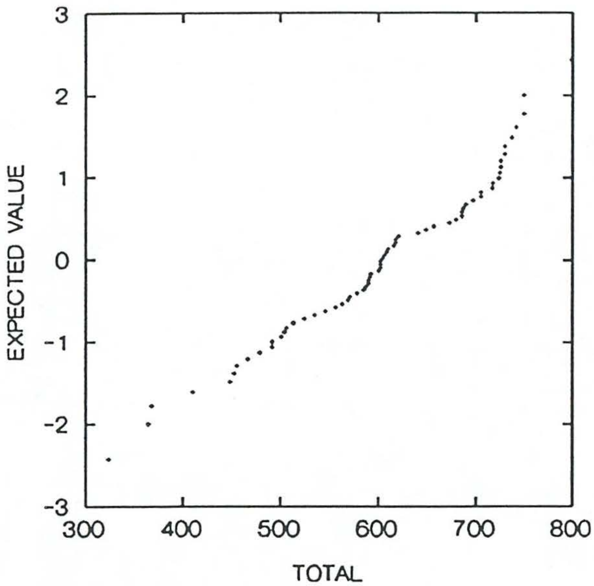


FIGURE 8

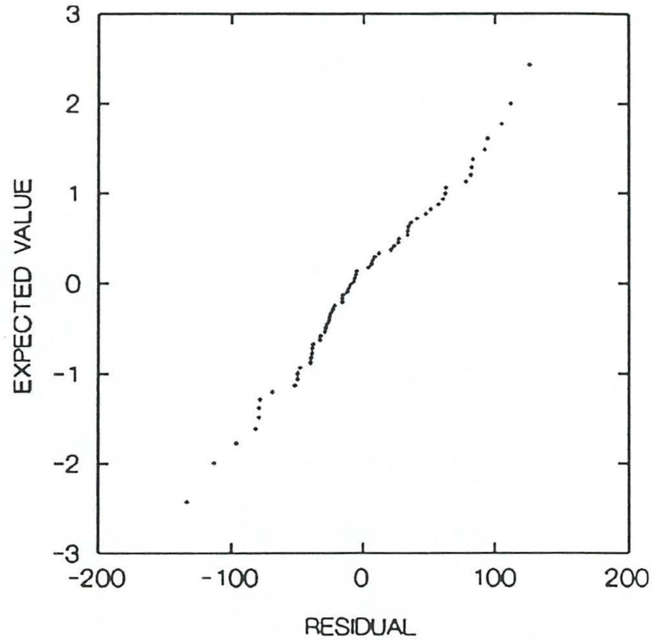


FIGURE 9

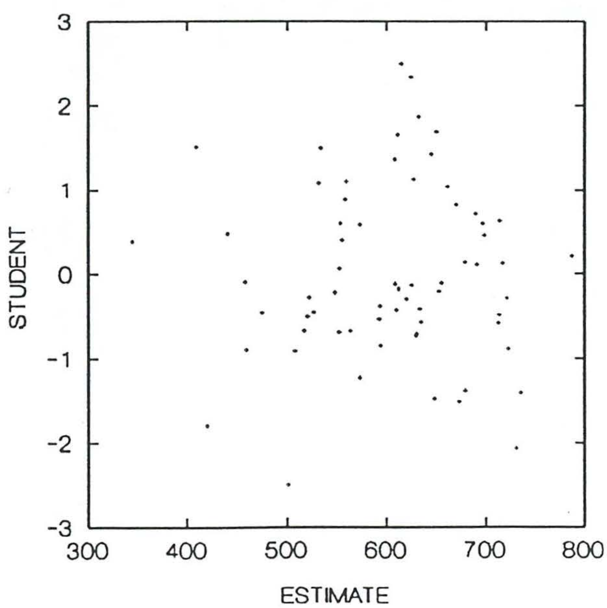


FIGURE 10