A GIS Developed Mapping Protocol to Determine Optimal Areas for Shoreline Restoration

Michael J. Goodnature
Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987

Keywords: Geographic Information Systems (GIS), Shore Zone, Grid, Raster, LIDAR, DEM, K-factor, Contingency Tables, Probability

Abstract

Like many urbanized lakes within the Twin Cities metro (MN) area, the shoreline of Long Lake has been highly developed for residential use. The removal of natural shoreline vegetation during residential development has allowed surface runoff to enter the lake without being filtered. Shoreline restoration efforts on Long Lake have been proposed, and a field inventory assessment of the shoreline was based on three factors: land cover, slope, and soil erodibility. An assessment was completed and incorporated into a model to identify areas in most need of restoration. These three factors were ranked and added together to determine which areas had low, medium, or high potential for the ability to effectively filter chemicals and sediment from runoff before it entered the lake. The areas classified as high were defined as areas with little or no potential to filter runoff, and were considered ideal for future restoration projects. This paper illustrates how a Geographic Information System (GIS) was used to develop a model to locate high priority areas using the same three factors listed above using Environmental Systems Research Institute's (ESRI) ArcGIS software suite. The goal in developing this model was to determine its accuracy and efficiency such that the protocol could be used for future models to save field data collection time and money. A raster layer created from the GIS model compared favorably to the raster created from the field data derived model. However, when the land cover and slope factors used in the GIS model were compared separately, there were not enough samples from each dataset to create an accurate comparison. The GIS model saved time and resources, but additional data may be needed for a more precise model.

Introduction

The Minnesota Department of Natural Resources (MN DNR) defines the shoreline impact zone as the area located between the ordinary high water mark of a public water and a line parallel to it setback 50 feet landward (MN DNR, 2007). A well established native vegetated shoreline within this zone plays a major role in filtering sediment and chemicals from entering the surface water of a lake. Altering a shoreline by removing natural vegetation for turf grass replacement or the building of structures can have detrimental effects on water quality, terrestrial and aquatic habitat, and increase the occurrence of shoreline erosion. The conventional ideal of most urban lakeshore is the expanse of turf grass mowed all the way to the water's edge, which has led to declining lakeshore habitat and water quality on many lakes (Henderson et al., 1998). The loss of shoreline vegetation in conjunction with erodible soils and steep slopes can increase erosion and the overland flow of chemicals and sediment.

Long Lake, located in the north metro suburb of New Brighton, MN, has experienced an increase in development and alteration within the shore zone in the past

Goodnature, Michael J. 2009. A GIS Developed Mapping Protocol to Determine Optimal Areas for Shoreline Restoration, Minnesota. Volume 11, Papers in Resource Analysis. 10 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. Retrieved (date) from http://www.gis.smumn.edu

60 years (Figure 1). A series of aerial photos dating back to 1940 depicts the increase of

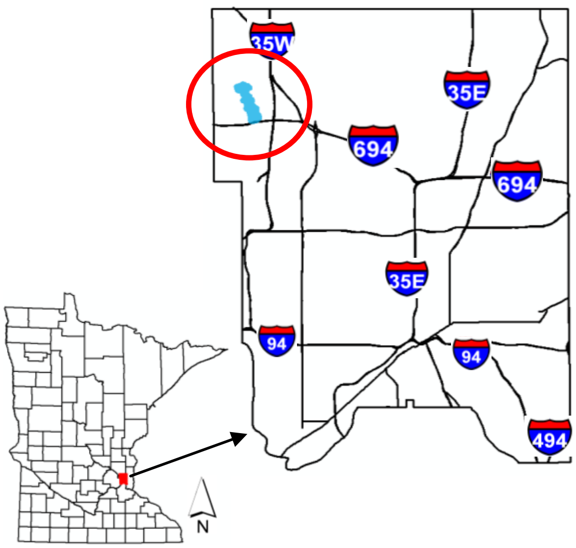

Figure 1. Ramsey County's location within Minnesota and Long Lake's location within Ramsey County.

development surrounding Long Lake (Figure 2). Resource managers believe that the increase of this development has explicitly led to the overall increase of shoreline erosion, low aquatic vegetation populations, and high phosphorous levels within Long Lake.

Existing survey information and lake assessment data shows that Long Lake has increased nutrient loading and low water clarity. Between the years of 1997 and 2006 data collected by the Minnesota Pollution Control Agency shows phosphorous levels in the lake average 103±4 parts per billion (ppb) for the northern basin of the lake and 55 (ppb) for the southern basin of the lake. When phosphorus concentrations exceed 40 or 50 ppb, they can produce algae blooms and create turbid water conditions which reduce water clarity. Algae are abundant in the lake, increasing turbidity, which limits aquatic macrophyte plant growth. A macrophyte plant survey was conducted in June of 2008 to inventory aquatic vegetation within Long Lake. Out of the 189 points surveyed within the littoral zone, 51 points contained aquatic plants (Ramsey Conservation District, 2008).

In late July of 2008, the Ramsey Conservation District (RCD) assessed the shoreline of Long Lake to determine the current shoreline quality with respect to shoreline erosion and presence of natural vegetative buffers. The assessment consisted of using a combination of data collected in the field and GIS layers to determine a rank

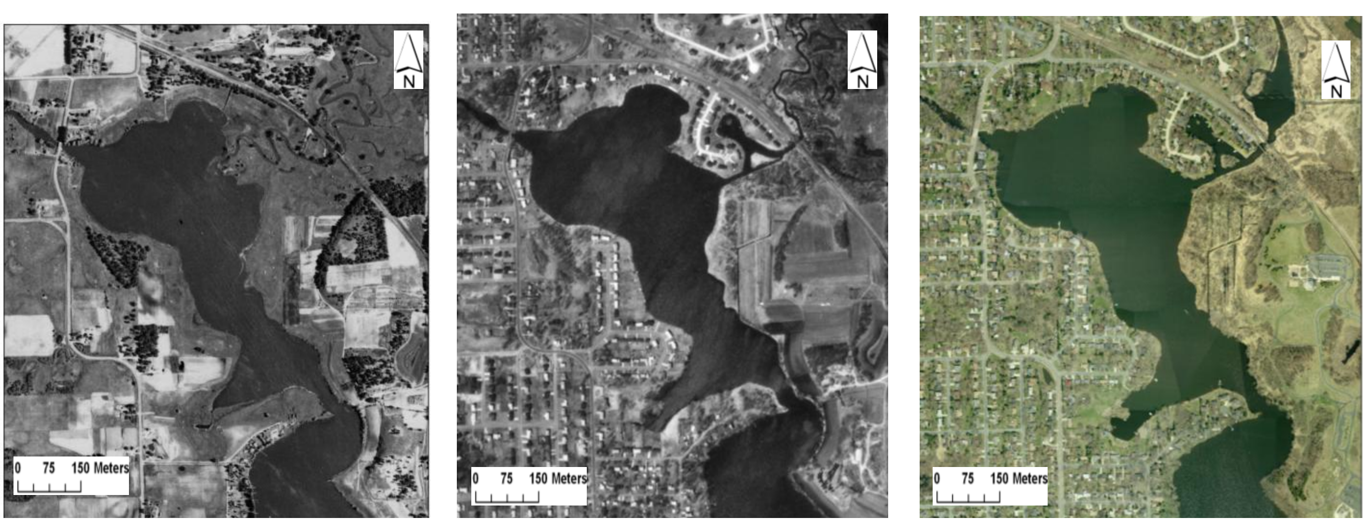

Figure 2. The increase in sprawl and development surrounding the northern half of Long Lake is shown in 1940, 1974, and 2008 (from left to right).

of buffering quality by assessing three variables: land cover, slope, and soil type, within the shore impact zone of each property parcel. Areas around the lake were identified as having high, medium, or low, potential for filtering surface runoff before it entered the lake. Implementing these categories allowed natural resource managers to determine which areas should be a priority in implementing grant funds for shoreline restoration projects.

Due to the size of the lake and the number of people needed to conduct the study the total time invested in field data collection alone was approximately 120 hours. With shoreline restoration costing anywhere from five to ten dollars a square foot, much of the financial resources used to collect the field data could have been used for cost sharing in the actual restoration of the shoreline.

This research project focused on using GIS software and datasets, as an alternative to using field derived data, to create a modeling process that could save time and money and still effectively identify high priority areas in need of restoration. To develop the model, the same criteria used to rank the field data was used to rank the GIS data. The ranking systems for the three variables (land cover, soil erodibility, and slope) were created using subjective standards developed by RCD personnel.

Methods

Development of the Slope Grid

The first step in creating the slope grid included processing Light Detection and Ranging (LIDAR) data to create a Digital Elevation Model (DEM). The bare earth LIDAR data consisted of a vertical accuracy of less than six inches with an average spacing of around 12 feet between elevation points. The LIDAR data were processed using a method developed by the MN DNR. Although a quantitative analysis of this method has not been established, the MN DNR has visually inspected and developed this method through years of experience (Loesch, 2008). A file geodatabase was created, and a feature dataset was added to contain the feature classes and LIDAR data to be processed. A subset of LIDAR points, housed within an ArcGIS SDE as a point feature class surrounding the lake were selected and imported into a feature dataset. An expanded boundary polygon and lake boundary polygon were also imported into a feature dataset. A terrain dataset, a TIN-based surface built from the measurements of the LIDAR points and boundary polygon feature classes, was then created within a feature dataset. The LIDAR subset was chosen to create the elevation for the TIN, at an average of a 12 foot distance between each point. The boundary polygon was chosen and used as a soft clip Surface Feature Type (SFType), and the lake boundary was chosen to create a hard value fill SFType. For quicker processing, three pyramid levels were created to allow for quicker drawing of the terrain model image when a smaller scale was chosen (Figure 3). The Terrain to Raster function was then used to convert the terrain to a one meter grid. The output data type chosen was "float," the interpolation method chosen was "linear," and the cell size created was 1 meter (3.2 feet). Using the focal statistics tool within the Neighborhood toolset in the Spatial Analyst toolbox, the one meter DEM was smoothed using a 3 by 3 rectangular cell window. To verify the accuracy of the DEM, two foot contours were generated using the surface analysis function and compared to Ramsey County's survey tested contours derived from raw LIDAR points (Figure 4). The two foot contours were reviewed by Ramsey County GIS personnel and were found to accurately depict the survey tested contours generated from raw LIDAR data.

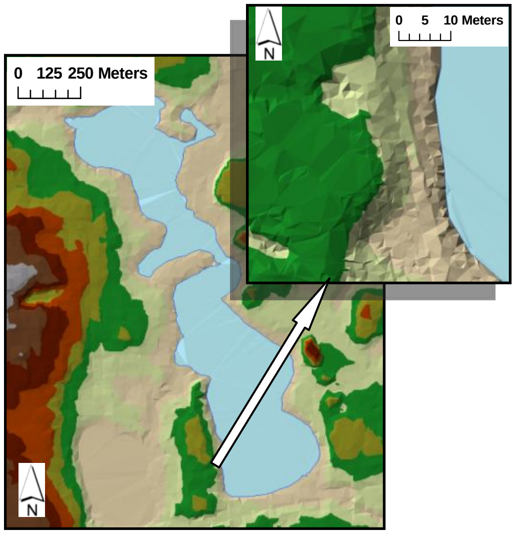

Figure 3. Terrain layer of the study area depicting resolution of 3rd pyramid level at 1:15,000 (lower left) compared to 1st pyramid level at 1:600 (upper right).

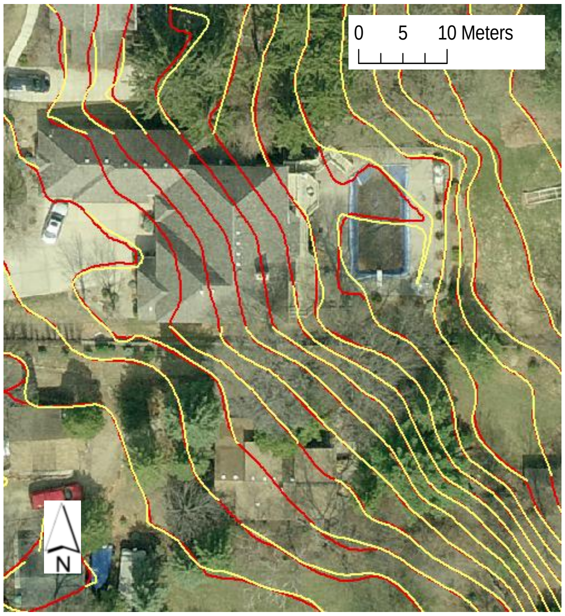

Figure 4. A comparison of Ramsey County contours (yellow) to contours derived from the terrain model (red).

Using the enhanced DEM to create two foot contours allowed for smoother lines and no gaps in the contour data that otherwise existed within the two foot contours derived from the raw LIDAR data.

Using Spatial Analyst, a percent rise slope grid was created from the smoothed DEM. The slope grid was reclassified where ≤10:1 (1/10*100) = ≤10 grid cell values, >10:1 (1/10*100) - 5:1 (1/5 *100) = >10-20 grid cell values, and >5:1(1/5 *100) – vertical = >20-90 grid cell values. The cell groups were then reclassified as follows: 0-10 = 1, 10-20 = 2 and 20-90 = 3 and contours were overlaid for comparison (Figure 5).

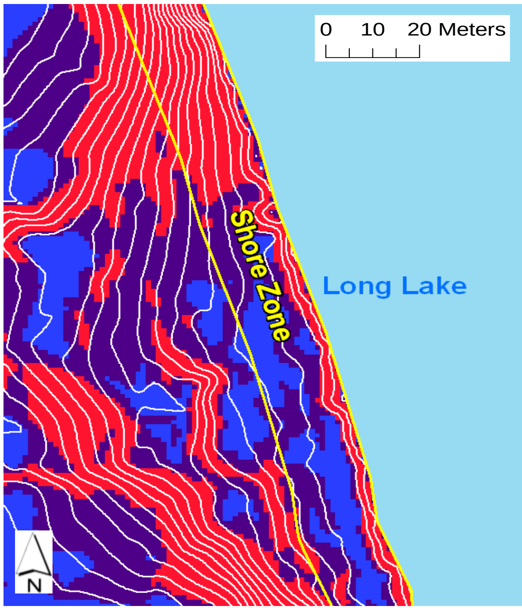

Figure 5. The reclassification of the slope grid with overlain contours for comparison. The slope ranks are depicted as Blue = 1(≤ 10:1 slope), Purple = 2 (>10:1 - 5:1 slope), and Red = 3(>5:1-Vertical slope).

The reclassified slope raster was then exported as a raster to a feature and intersected with the parcels layer so that the dominant slope rank could be identified within each parcel. A field was added to the slope layer and total square meters were calculated for each rank within the corresponding parcel. The slope field was then imported into Microsoft Access where a query was run to determine which classification of 1 (0-10), 2 (10-20) or 3 (20-90) was dominant within each parcel. The rank representing the slope for each parcel was selected and added to a final slope rank field. The slope feature was then converted back to a raster so that it could be added to the soils and land cover grid.

Development of the Land Cover Grid

The land cover data was created using parcel boundaries and six inch resolution aerial imagery of the lake and surrounding area. The parcel boundaries were imported into a file geodatabase containing coded value domains, which represented the land cover rank values, of 1 = Native Vegetation, 2 = Weedy Vegetation, 3 = Turf Grass, 4 = Bare Soil, and 5 = Impervious Surface. The land cover was ranked from lowest to highest depending on the cover type which had the most to least potential to filter runoff. At a scale no larger than 1:300 the aerial imagery was interpreted and a dominant land cover was chosen for each parcel. The land cover layer was then exported from a feature to a raster using the rank classification as the grid cell ID.

Development of the Soils Grid

To determine the soil types within each parcel surrounding the lake, the NRCS Soil Survey Geographic Database (SSURGO) spatial and tabular data for Ramsey County were downloaded and appended together. The soils layer was then intersected with the parcel layer and a soils rank field was added to the attribute table. This field was populated using the following reclassification schema: sand (K-Factor ≤ 0.2) = 3, loam (K-factor 0.2 < 0.28) = 2, or clay (K-factor ≥ 0.28) = 1. The higher the K-factor, the greater the soils type was prone to erosion. The soils layer was then exported from a feature to a raster using the rank classification as the grid ID. There were seven different soil types that surrounded Long Lake. The majority of the soils had a K-factor of 0.17, which classified these areas with a rank of three. A small area of soils located in the northeast corner surrounding the lake was made up of Chaska silt loam, and therefore had a K-factor of 0.28. The areas that fell within this soil type were ranked as a two.

Grid Overlay

To create the final raster layer, the soils, land cover, and slope grids were added together and reclassified. The three layers were added together using the Raster Calculator to produce a grid with values ranging from 5-10. These values were then reclassified using the Raster Calculator to the following: 5-6 = 1 (low), 7-8 = 2 (medium) and 9-10 = 3 (high).

To determine if the model located the same high priority areas as the field derived model, the Spatial Analyst extension was used to reclassify and overlay the two grids. Grid values of 1 and 2 within the field model were classified as zero and values of 3 were classified as 1. In the GIS model, grid values of 1 and 2 were also classified as zero, but values of 3 were classified as 2. The values were classified in this schema so that when the two grids were added together the final grid would consist of 0, 1, 2, or 3 cell values. This was used to determine if the models matched (cell values 0 or 3) or if the GIS model (cell value 2) did not correspond with the Field model (cell value 1), or vice versa.

The two grids were added together to create a final grid in which low/medium areas consisted of a shore zone area capable of filtering runoff and high areas, which consisted of a shore zone that provided little buffering potential for the filtering of runoff.

The final grid consisted of the following classification: 0 = areas of low/medium priority for both models, 1 = areas of low/medium in the GIS model and high in the field model, 2 = areas of low/medium in the field model and high in the GIS model, and 3 = high priority areas in both models (Figure 6). This raster was then

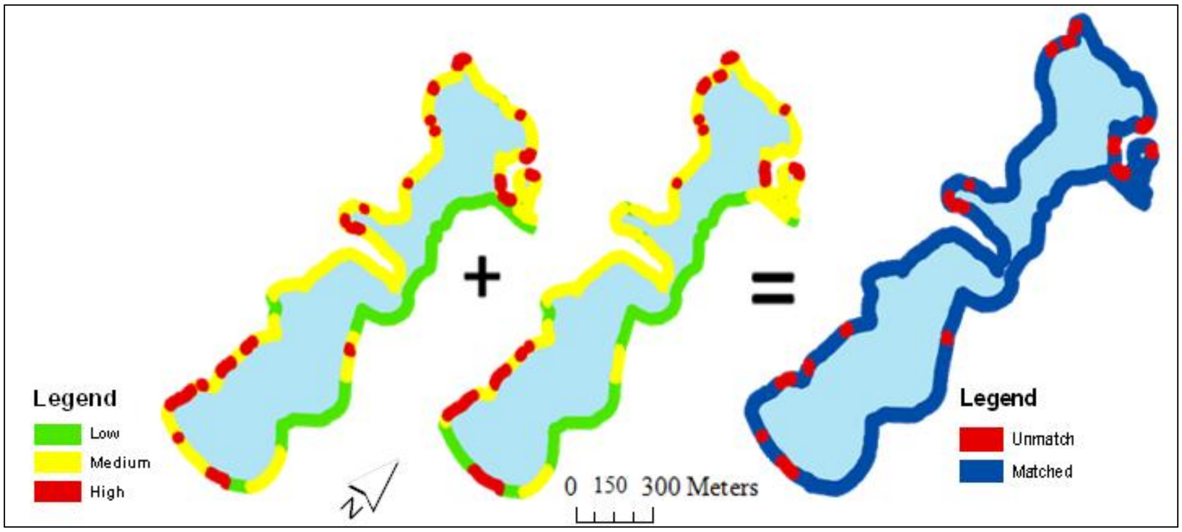

Figure 6. The field model (left) was added to the GIS model (center) to create the final grid model (right), depicting areas where the two models did not match.

converted to a feature layer so that the difference in areas could be calculated.

To examine the variables within the two models, contingency tables were created for both the land cover and slope variables (Table 1) so tests could be conducted to determine relationships between variables within the models. The numbers within the contingency tables consisted of frequency of observations in each category for land cover and slope. For example, there were 96 predicted cases of weedy vegetation within the GIS model that matched the observed data within the field model. The data shown in the contingency tables were entered into the statistics program JMP 7. Within this program, the predicted (GIS data) were entered into the Y axis and the observed (field data) were entered into the X axis. A nominal logistic model was used to analyze the paired observed data (field data) to the expected data (GIS data) for the land cover and slope variables separately. The nominal logistics test was chosen in the personality field and the model was run. The model generated the chi-square statistic and error probability between the predicted and observed datasets.

Table 1. Contingency tables for slope and land cover variables showing the frequencies of factors within each model. The soils data for both models were taken from the same layer so there was no variation for comparison.

LAND COVER		Observed				
		1	2	3	4	5
Predicted	1	0	0	0	0	0
	2	0	96	4	2	0
	3	0	24	125	2	1
	4	0	0	0	7	0
	5	0	10	0	0	7

SLOPE		Observed		
		1	2	3
Predicted	1	105	38	8
	2	5	17	4
	3	4	11	14

Results

The overall findings within the shore zone found that of the approximately 115,215 total square meters surveyed, 37,397 were classified as low by the GIS model, 66,617 were classified as medium and 11,202 were classified as high. A comparison of these numbers to the field data model findings is illustrated in Table 2. Of the 115,216 total square meters surveyed, 7,850 square meters did not coincide between the two grids. Table 3 illustrates the difference between the reclassified numbers. Grid codes 0 and 3 were a match between the two grids and values 1 and 2 represented the differences calculated between the two grids.

Table 2. The difference is shown in square meters of the low, medium, and high classified areas between the two models.

Areas with runoff problems	*GIS Model*	*Field Model*
Low	37,397	33,858
Medium	66,617	70,683
High	11,202	10,675

Table 3. The areas that coincided (0, 3) and the areas that did not match (1, 2) between the two models.

Grid Code	*Square Meters*	*Percent*
0,3	107686	93%
1,2	7850	7%

The areas of difference were examined further to determine which factors (land cover, slope, or both) were involved in creating a variation between the two grids (Table 4). The soils layer did not need to be compared since the GIS layer used to create the soils rank was the same for both the field and GIS models. The results generated from the JMP 7 software showed a P-value of <.001 for both the slope and land cover data tested. The chi-square statistic for the land cover and slope data tests were 282 and 61 respectively. The degrees of freedom, a measure of the independent information used in the calculation, was calculated as follows (Gotelli and Ellsion, 2004):

df = υ = (number of rows -1) x (number of columns -1).

This resulted in degrees of freedom of 4 for the slope data and a 9 for the land cover data. Mosaic plots were also generated

Table 4. The area, shown in square meters, of the factors that did not coincide between the two models is shown below.

Grid Difference	*Sq Meters*
Slope	5061
Slope/Land Cover	2427
Land Cover	362
Total	7850

from the JMP 7 software. The mosaic plots depicted values in the contingency table's cells as tiles to better visualize the datasets (Figure 7). The illustrations show the relationship between the observed (field derived) data (X-axis) to the predicted (GIS derived) data (Y-axis). The tiles sizes are proportional to the frequency in the dataset (Gotelli and Ellsion, 2004). For example, within the land cover table (top), the red tile, with horizontal axis label "2", is a visual representation of how many incidences weedy vegetation was recorded within each parcel in the field (observed axis) and also recorded within that same parcel using aerial photo interpretation for the GIS model (predicted axis). The green tile in column "2" represents how many times weedy vegetation was observed in the field, but mistaken for turf grass during aerial photo interpretation for the GIS land cover grid. The brown tile in column "2" represents the frequency of observations where weedy vegetation was observed in the field, but was mistaken for impervious surface during aerial photo interpretation. Similarly, column "3" corresponds to turf grass; column "4" to bare soil; and column "5" to impervious surface (brown tile).

Similar to the land cover table the slope table (bottom) depicts the frequency in red where ≤ 10:1 slope (value 1 along the observed and predicted axis), was recorded both out in the field and in the LIDAR derived grid, within the same spatial location. Concurrently, >10:1 - 5:1 (value 2) is represented as green, and >5:1 – vertical (value 3) is represented by blue to represent the incidences between the two grids.

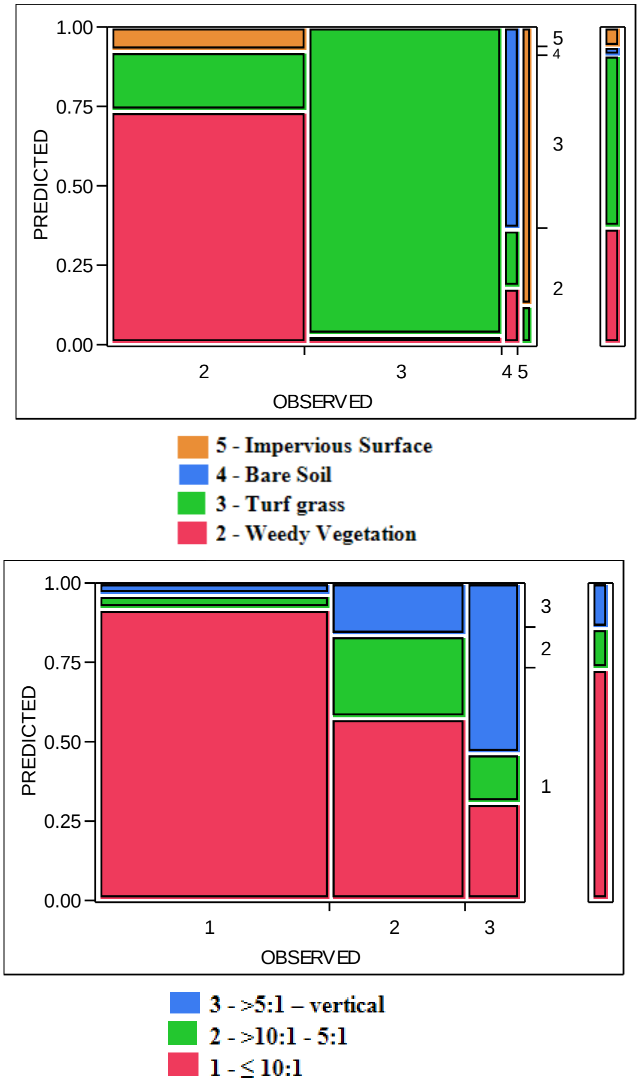

Figure 7. Contingency analysis of predicted (Y-GIS model) by observed (X-field model) mosaic plots for the land cover (top) and slope (bottom) variables.

Discussion

With 93% of the GIS model matching the field derived model it can be suggested that the methods used to create the GIS model could be used to determine future shorelines in need of restoration. However, when the variables were reclassified and lumped into new categories of low, medium, and high, the association between the data results may not have been the same as if each variable of land cover and slope were compared separately (Reynolds, 1984). To ensure there was a relationship between the two variables, the chi-square test for nominal data was utilized and produced results that revealed land cover and slope were closely related to the observed field data. The contingency tables used to run these analyses were used to test the null hypothesis that the variables were not associated with each other (Gotelli and Ellsion, 2004).

The null hypothesis that there was no relationship between the paired observed data (field data) to the expected data (GIS data) for the land cover and slope variables was rejected. Both of the models yielded chi-square statistics that were highly significant ($P<.0001$), meaning the observed and predicted values were closely related. Although the P-value was highly significant for both datasets, the accuracy may have been skewed because more than 20% of the predicted versus observed table cells had counts less than 5. Because more than 20% of the table cells had counts less than 5, there were not enough observations within each category, which could make the chi-square test statistic suspect. To create a stronger relationship between the datasets the number of observations should be increased such that the number of cells in the predicted versus observed table with counts less than 5 occurs in less than 20% of all the cells. Ideally, collecting more data from more diverse areas would increase the number of observations between datasets such that the less than 20% rule for application of the chi-square statistic is met. To increase the number of observations, it is suggested that more field and GIS data be inventoried on different lakes and added to the model for comparison.

The overall goal of this project was to develop a time and cost efficient method used to identify areas of concern surrounding bodies of water surrounded by development. The idea behind this model was to save funding by reducing field data collection so money saved could be invested into implementing restoration projects within high priority areas. With time invested in field data collection totaling 120 hours versus the time to go through the procedures to develop the GIS model (approximately 40 hours), the process proved to be a time and cost efficient measure.

Conclusion

This research outlined the methods used to create a GIS based model to identify areas prone to little or no buffering potential for filtering runoff entering Long Lake. The three variables (land cover, slope, and soils) used to create the model were ranked according to factors that could lead to the increase of runoff. Although this study was able to locate areas that were susceptible to accelerated runoff and erosion by combining the factors of steep slope, erodible soils, and lack of stable vegetation, the accuracy of the model developed is questionable. Although overlaying the field and GIS derived grids accounted for accurate results, assessing the variables that produced these models independently proved to have a strong relationship only if the number of cells within the observed versus predicted tables, with counts less than 5, could be overlooked. The limiting number of observations within the comparative contingency tables may have led to inaccurate P-values amongst the datasets. If future data is added into the tables, increasing the number of observations, the chi-square statistics obtained for the model predictions can be more confidently interpreted. The time invested in developing the procedure used to create the model was minimal compared to the time invested to collect and create data from the field. If additional data is incorporated into the model and proven to have a strong relationship, resource managers may be able to use this procedure to apply the model to other highly developed lakes within the Twin Cities metro area, saving time and resources.

Acknowledgements

I would like to thank Saint Mary's Professors Dr. McConville, John Ebert, and Patrick Thorsell, who have been an inspiration for the advancement of higher learning amongst their students. I would also like to thank my family, friends, and especially Ali for her constant support and insight. I would also like to recognize the Ramsey Conservation District for whom a portion of this work was completed.

References

Gotelli, N. J., and Ellsion, A. M. 2004. A Primer of Ecological Statistics. Sinauer Associates Inc., 23 Plumtree Road, Sunderland, Massachusetts, USA. 447 pp. www.sinauer.com.

Henderson, C., Dindorf, C., and Rozumalski, F. 1998. Lakescaping for Wildlife and Water Quality. Nongame Wildlife Program Section of Wildlife, Minnesota Department of Natural Resources, 500 Lafayette Road, Saint Paul, Minnesota, USA. 175 pp.

Loesch, Timothy N. 2008. From LAS to Contours. Workshop Presentation. Workshop conducted at the Minnesota GIS/LIS Fall Conference, Rochester, Minnesota, USA. http://www.mngislis.org/displaycommon.cfm?an=1&subarticlenbr=436.

Minnesota Department of Natural Resources (MN DNR). 2007. Minnesota's Sensitive Lakeshore Area Identification Manual, a comprehensive wildlife conservation strategy for Minnesota lakeshores (version

0.9). Division of Ecological Services, Minnesota Department of Natural Resources, 500 Lafayette Road, Saint Paul, Minnesota, USA. 64 pp.

Ramsey Conservation District. 2008. Long Lake Macrophyte Survey. Ramsey Conservation District and Ramsey County Public Works Lake Management, 1425 Paul Kirkwold Drive, Arden Hills, Minnesota, USA.

Reynolds, H. T. 1984. Analysis of Nominal Data. Second Edition. Sage Publications, Inc., Newbury Park, California, USA. 85 pp. order@sagepub.com.