

Geography 586

Watershed Analysis in New York State

Submitted by
Ruth Ann Trudell
12/21/2007

Introduction

The goal of this project is to assess the relationship between stream water quality and watershed population, land use, slope, road density and toxic substance releases. Like elsewhere in the country, maintenance and improvement of water quality enjoys considerable interest in the State of New York. Also, as with other states, dollars to remediate problems are limited. If a strong correlation is found to exist between stressors and water quality, it may be possible to design preclusive measures to address potential impacts before they develop into costly problems.

The objectives of the project are threefold: 1. determine if a ranking system for stressors can be used to predict water quality outcomes, 2. determine if a correlation exists between specific stressors and water quality and 3. determine if underlying spatial autocorrelation is associated with the magnitude of any of the analyzed stressors.

Ranking of watersheds relative to each other according to the magnitude of selected stressors present has been used by federal agencies to identify watersheds in need of protection or improvement measures. The ranks represent the relative potential magnitude of selected stressors and their potential impact on water quality. This predictive approach will be assessed against actual water quality outcomes represented as impairment classes determined by the State of New York,

Correlation values will be determined for each of the stressors indicating their significance in determining water quality outcomes. If indications are that those correlations exist, the data set can then be broadened to further test the hypothesis and refine the measures used. Follow-up could include testing and potentially expanding the list of stressors leading to the development of a comprehensive approach to watershed protection.

If spatial autocorrelation is found to be operational, this may help focus mitigation measures to specific geographic regions.

Methods

Fortunately, the Department of Environmental Quality for the State of New York has stored relatively recent versions of all data necessary for the completion of this project on its internal server. These data layers include:

- a. 11 digit HUC's (Hydrologic Unit Code)- These boundaries were delineated and published by the Natural Resources Conservation Service at the 1:24000 scale. Version: December 30, 2004.
- b. Classified stream segments- This dataset provides a summary of general water quality conditions, tracks the degree to which a waterbody supports its designated uses, and monitors progress toward the identification and resolution of water quality problems, pollutants, and sources. Version 2006.
- c. Population data by census tract- A small, relatively permanent statistical subdivision of a county delineated by a local committee of census data users for the purpose of presenting data. Census tract boundaries normally follow visible features, but may follow governmental unit boundaries and other non-visible features in some instances; they always nest within counties. Designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time of establishment, census tracts average about 4,000 inhabitants. Version 01/01/2003.
- d. Toxic Release Inventory- This dataset shows the locations of Toxic Release Inventory (TRI) facilities in NYS. The TRI is a public database of annually reported toxic chemical releases and management from manufacturing or processing facilities that produce, process, or otherwise use a TRI chemical in excess of a certain threshold level. These facilities must report TRI information to the U.S. Environmental Protection Agency (EPA) and to the state in which the facility is located. The inventory was established to provide information to the public about the presence and release of toxic and hazardous chemicals in communities. Version: 1997
- e. NY State digital elevation matrix (DEM)- Based on 10 meter DEM obtained from the US Geological Survey.
- g. NY State ALIS road layer- The Accident Location Information System (ALIS) project is a multi-agency project that the NYS Office of Cyber Security & Critical Infrastructure Coordination (CSCIC) is jointly developing with the NYS Department of Motor Vehicles (DMV) and the NYS Department of Transportation (DOT). A major component of the ALIS Project is the creation of an up-to-date statewide Geographic Information System (GIS) street map file containing all public roads, along with their street names, alternate/alias street names, route numbers, and address ranges on each street segment. Version 08/2005.
- h. Land Use Land Cover- obtained by the State from the USGS. Version 1985-1997.

Due to time constraints, only a limited number of watersheds were examined. This was felt to be appropriate due to the testing nature of the project. The selected watersheds included those where there is clearly an abundance of impaired streams and those where impaired channels are clearly lacking. Using the subset of watersheds allowed for testing and fine tuning the evaluation processes; choosing watersheds that were clearly impaired and unimpaired avoids the issue of the dilution of the cause effect relationships with streams in the watersheds that are either not sampled (perhaps they are actually impaired or unimpaired) or have an impairment status opposite the majority of the streams in the given watershed.

The project watersheds are shown in Figure 1 below.

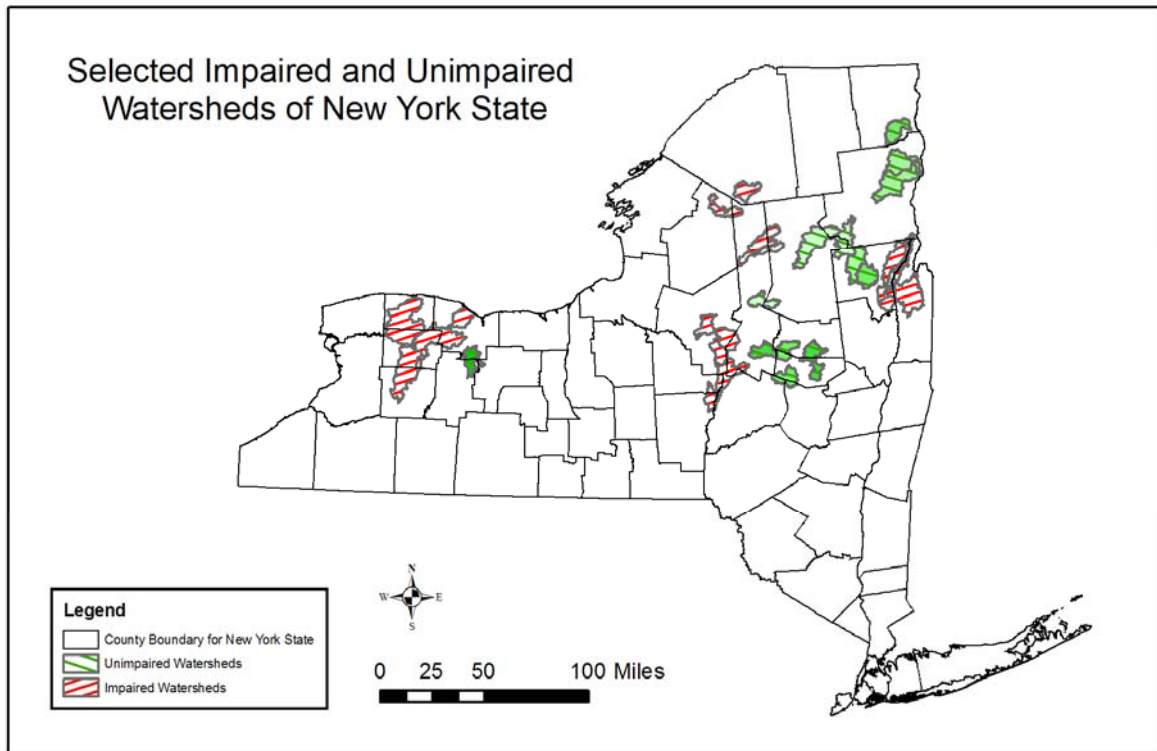


Figure 1. Map showing selected impaired and unimpaired watersheds for the state of New York.
Source: all data provided by NY State DEC intranet web server, mapped using ESRI ArcGIS 9.2 software.

The following steps were taken to process the data. All data were normalized for comparison purposes.

1. In order to develop the final unimpaired and impaired watersheds layer for the study, the entire watershed data set was screened for watersheds that contained only 1) impaired and non-sampled streams and 2) unimpaired and non-sampled streams. From the screening only those watersheds with a preponderance of streams in the impaired or unimpaired categories (>80%) were selected. As a result of the screening process, 28 watersheds were finalized (see map above). These watersheds range in size from 34 to 272 square miles.
2. The percentage of land in slopes greater than 10 degrees by watershed was extracted from the DEM data.
3. A population surface was generated by first creating centroids from census tract polygons and then interpolating between the points using the Inverted Distance Weighting and kernel density tools in ArcGIS. The zonal statistics function was used to sum the number of residents per grid calculated for the density map by watershed. This process step was taken because the census tract and watershed boundaries are not coincidental and therefore this technique would result in the most accurate representation of population calculated across the watershed landscapes. Finally the population for each watershed was normalized to # of individuals per square kilometer by watershed.

4. Toxic release sites were counted and normalized as a density per kilometer (km), thereby allowing for comparisons between watersheds of differing sizes.
5. Land Use Land Cover- The *Tabulate Area* function was used to extract the number of cells with values of Agriculture, Industrial, Residential and Mining were summed per watershed and normalized as a percentage of land area in the watershed.
6. Roads were calculated as a density (km per square km) by watershed.

Once the data had been normalized, the watersheds were ranked according to the magnitude of each stressor in the watershed and the cumulative amount of stressors in each of the watersheds. Data were also analyzed using Pearson correlation to determine the probability that the stressor resulted in impairment. Finally the data was reviewed for clustering.

Results

The attached table (see link below) is the result of massaging the data listed and ranking the results. A rank of 1 indicates the least amount of stressors operational in a watershed, while a rank of 28 (28 total watersheds in study) indicates the highest magnitude of stressor. These rankings were performed for each of the individual stressors. Once ranks were determined for each of the individual stressors, they were then summed and the watersheds were again ranked by the sums. This process was followed in order to determine if the impairment status of the watersheds could be inferred from the overall final ranks.

The following link routes the reader to the data document ([WatershedLink.pdf](#))

Looking at the attached table, the rank columns represent the ranking of the watersheds from lowest to highest, with the lowest rank representing the lowest magnitude of stressor. For instance, the watershed with the highest population density (7956/ sq.km,) has also the highest rank (28). Where the value being ranked is equal between watersheds, the midpoint value is used as the rank for all. The last two columns in the table represent the sum of the individual parameter ranks and a re-ranking of the watersheds based on those sums respectively. In the end, the Rank of Sums column represents the integration of all stressors in the watershed. The map shown in Figure 2 was developed from the data in the table.

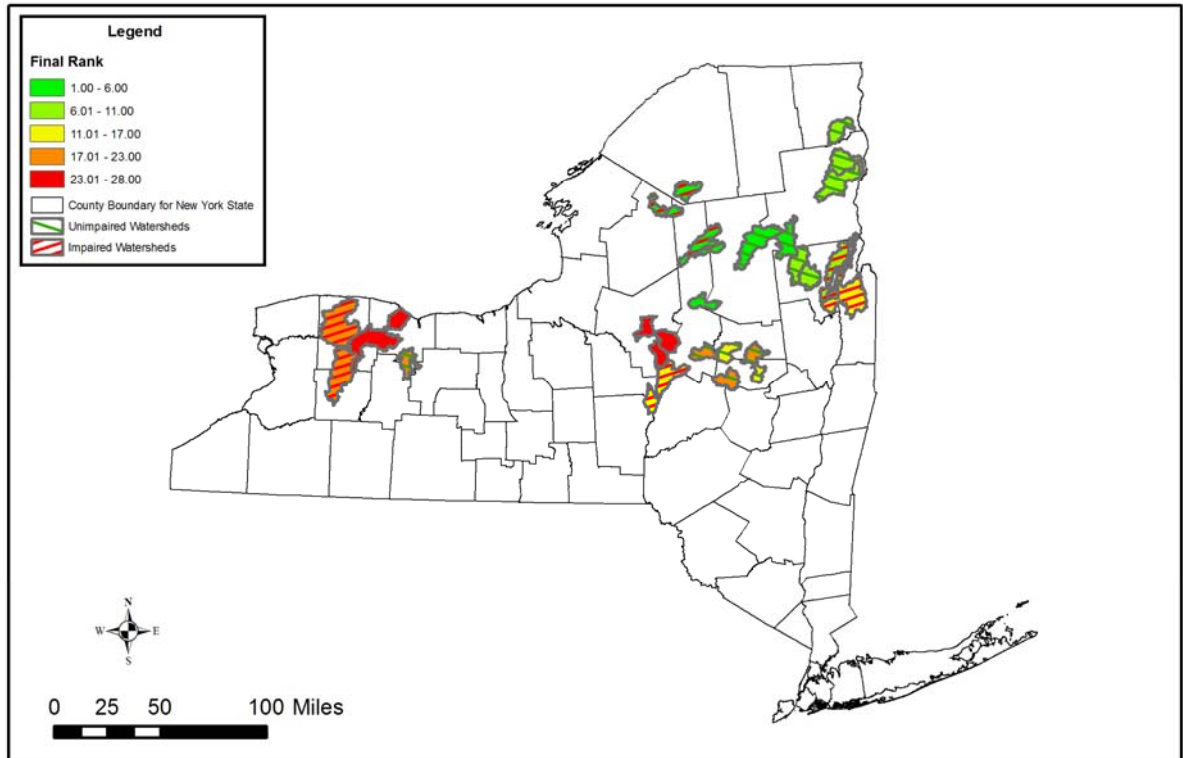


Figure 2. Map showing selected study watersheds colored illustrating the ranking of the watersheds based on stressors, red indicating a high overall level or concentration and green indicating low concentrations. Source: all data provided by NY State DEC intranet web server, mapped using ESRI ArcGIS 9.2 software.

What is apparent when comparing Figures 1 and 2 is that there is some consistency between rank and impairment, but there is also some discordance; e.g. note the red hash marks over green background in the north-west watersheds. This leads to the conclusion that further exploration and refinement of the data is needed. In deed additional paramenters not yet evaluated may be operational.

In addition to ranking the values of the stressors by watershed, probability calculations using Pearson correlation were also performed in order to determine the probability that the parameter is a cause of impairment. The results are as follows:

	<u>Probability</u>	<u>Correlation Value (r)</u>
Slope	-.3845	0.0434
Population	0.0953	0.6296
TRI	0.0272	0.8905
Non-Point	0.2389	0.2209
Road	0.2069	0.2909

As can be seen from these results, the strongest correlation was with the TRI values; strong enough to infer causality. The rest of the data showed no strong correlation,

although population density did show a weak relationship. Interestingly, slopes show an inverse relationship with impairment, that is, the more representation of higher slopes in the watershed, the less probability of impairment. One reason for the low values associated with roads and non-point sources may be their representation in watersheds that are NOT impaired.

The following maps are of the individual normalized parameter values by watershed. Please refer to Figure 1 when making a comparison of the parameter maps with impairment class.

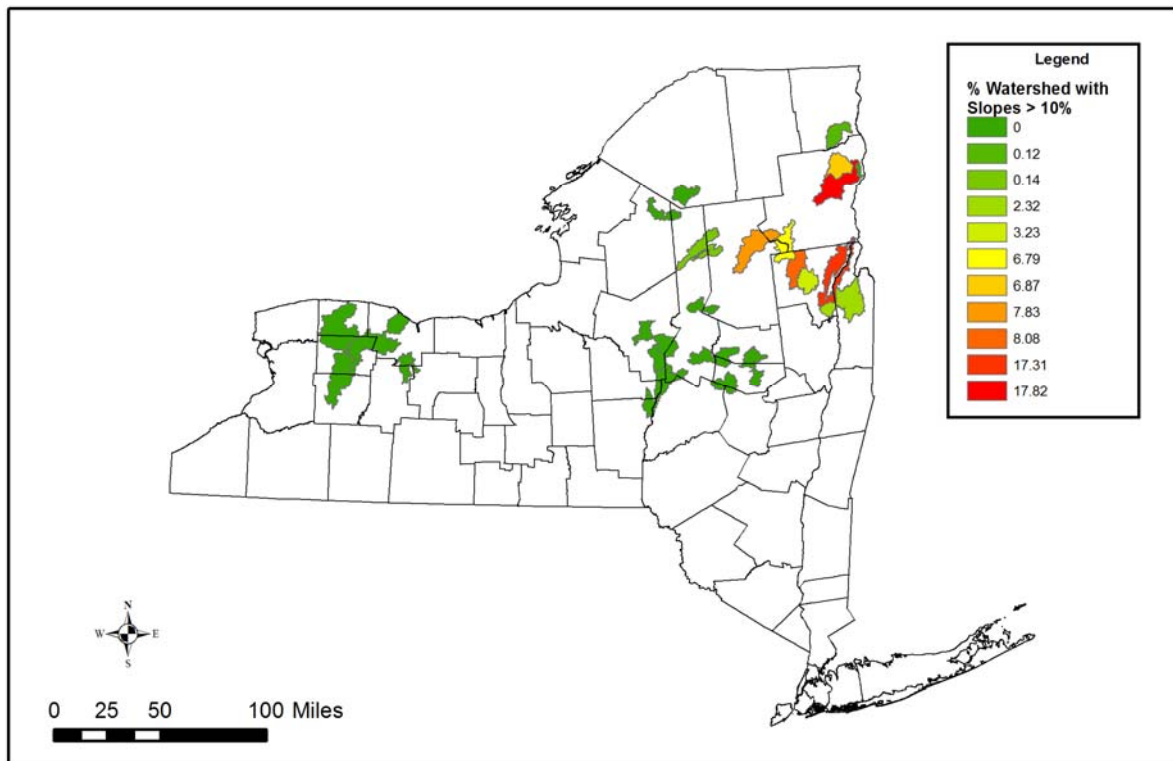


Figure 3. Map showing population the percentage of area in each watershed that is greater than 10%. Source: all data provided by NY State DEC intranet web server, mapped using ESRI ArcGIS 9.2 software.

Figure 3 illustrates that there is no correlation between the higher slopes and the impaired classification. This may be because sediment is not considered or is not weighed heavily in the impairment classification process. The areas indicating higher slopes are located in the vicinity of the Adirondack Mountains, where few roads, people, dump sites, and agriculture exist.

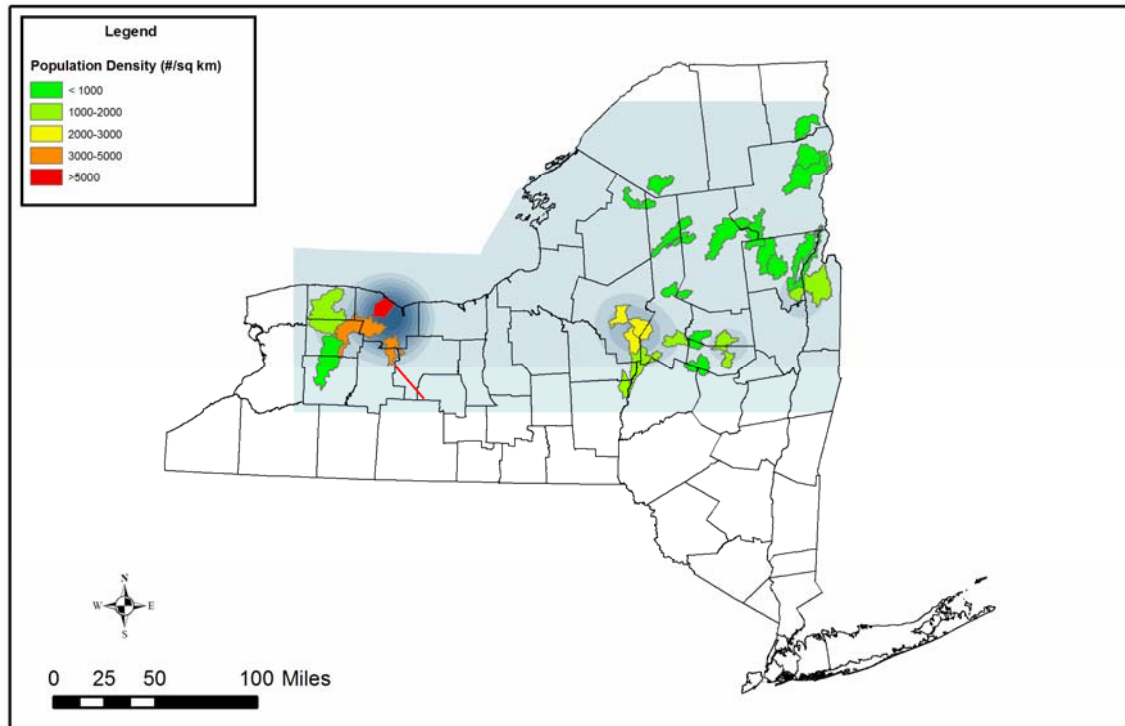


Figure 4. Map showing population density by selected watershed. Also note that the concentric blue circles represent the density surface that was created using the kernel density function. Source: all data provided by NY State DEC intranet web server, mapped using ESRI ArcGIS 9.2 software.

In Figure 4 above, it is clear that all but one watershed with higher densities of populations, (identified with the red arrow), appear to be impaired, but lack of population did not imply non-impairment. The high population are in the west also happens to be an area of intense agriculture.

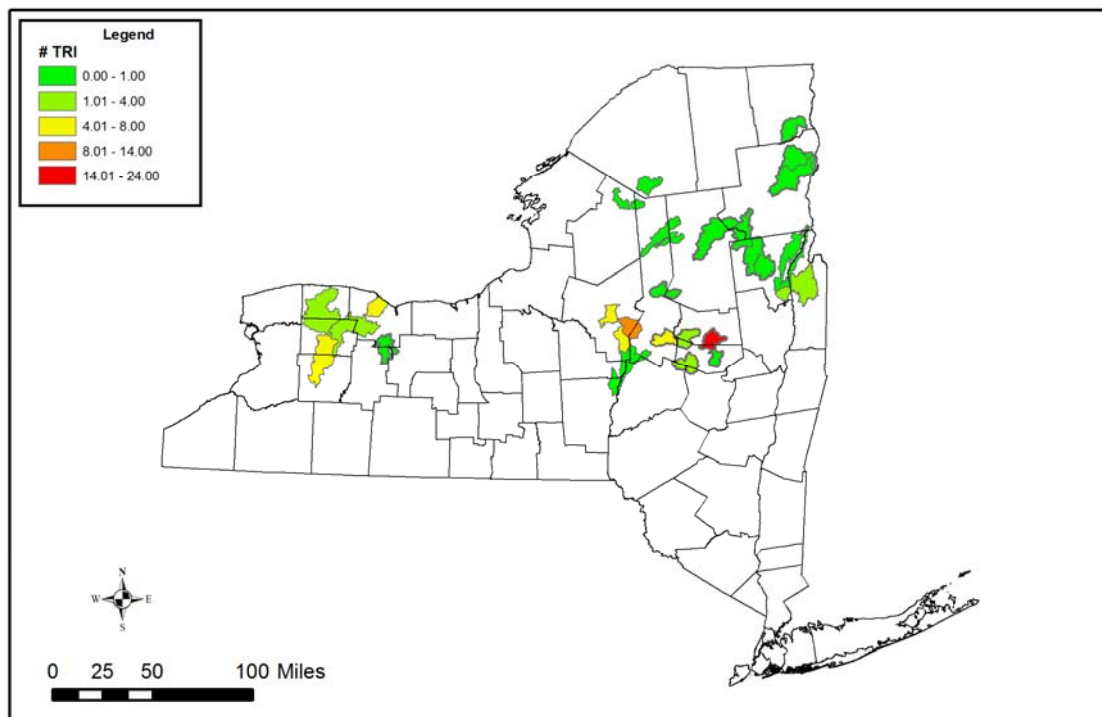


Figure 5. Map showing toxic releases by selected watershed. Source: all data provided by NY State DEC intranet web server, mapped using ESRI ArcGIS 9.2 software.

The number of toxic releases in a watershed correlated most closely with impairment classification, although, interestingly, the watershed with the most TRI sites was not listed as impaired. This may indicate a need to research the circumstances of release, the type of release and the monitoring done.

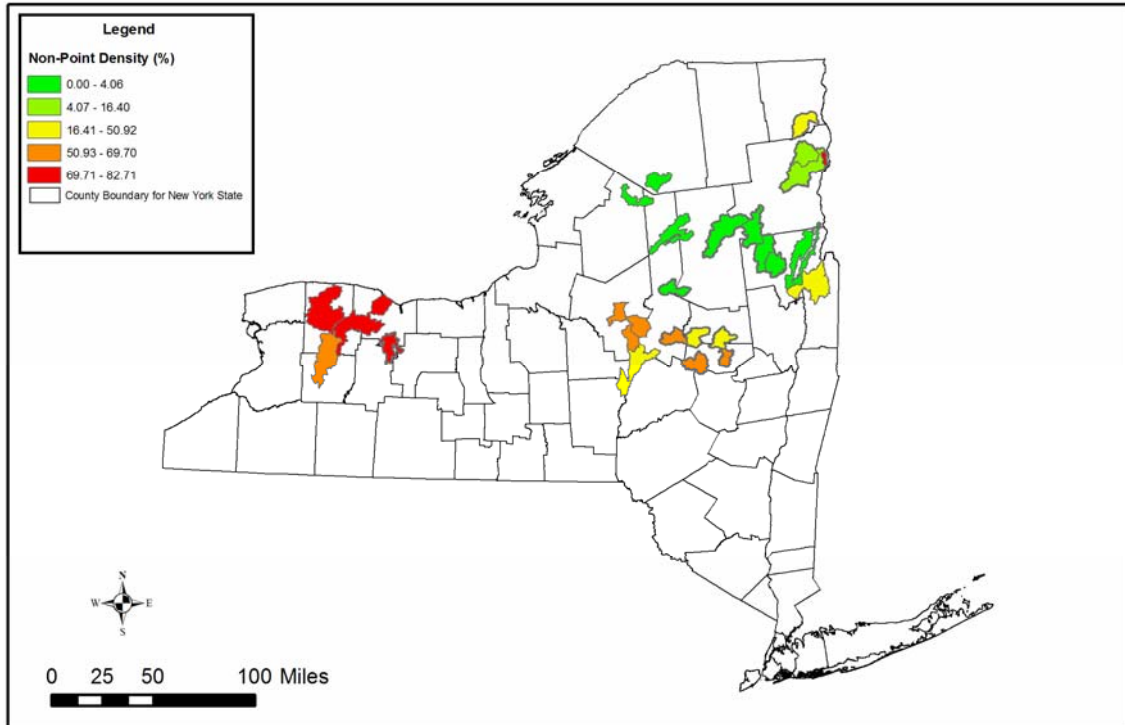


Figure 6. Map showing non-point pollutant generating activities by selected watershed. Source: all data provided by NY State DEC intranet web server, mapped using ESRI ArcGIS 9.2 software.

Geographically, the data presented in Figure 6 appears to make sense. Here it appears that most of the watersheds with high concentrations of non-point pollution sources are also impaired watersheds. There are also watersheds with higher concentrations of non-point activities that are NOT impaired.

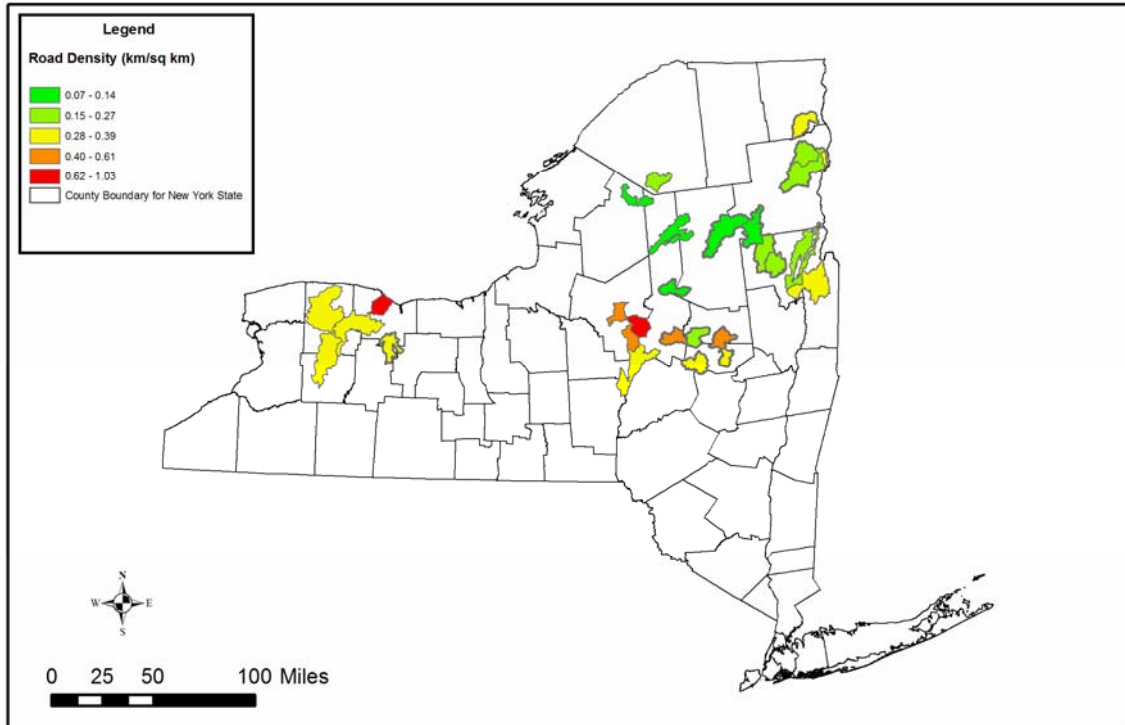


Figure 7. Map showing road density by selected watershed. Source: all data provided by NY State DEC intranet web server, mapped using ESRI ArcGIS 9.2 software.

The map above illustrates the loose correlation between road density and impairment class (not the middle of the state).

Finally, an analysis was done on spatial autocorrelation characteristics of the data. The following charts illustrate the results.

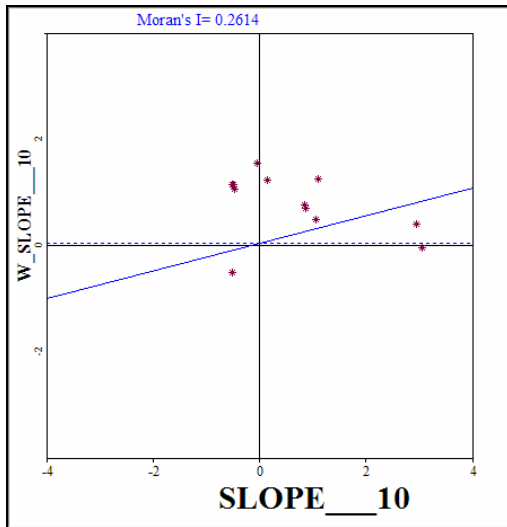


Figure 8. Moran's I value for watershed slope. The graph shows the percent area of a specific watershed in over 10 percent slope (X) versus other study watershed values weighted by distance (Y). The Moran's I = 0.2614 indicates little local spatial auto-correlation. Source: Chart developed using Geoda v 0.9.5 - i software.

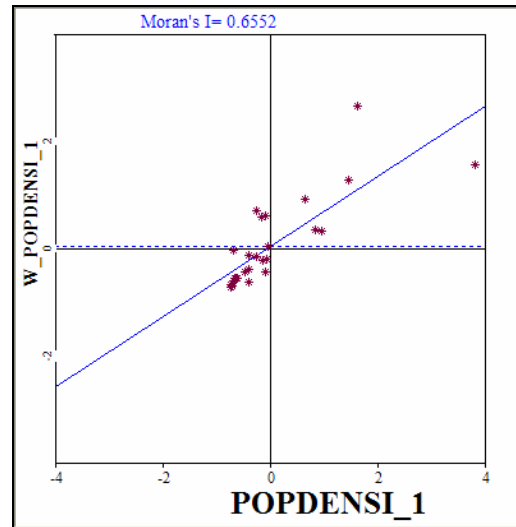


Figure 9. Moran's I value for population density. The graph shows population density in a specific watershed (X) versus other study watershed values weighted by distance (Y). The Moran's I = 0.6552 indicates local spatial auto-correlation. Chart developed using Geoda v 0.9.5 - i software.

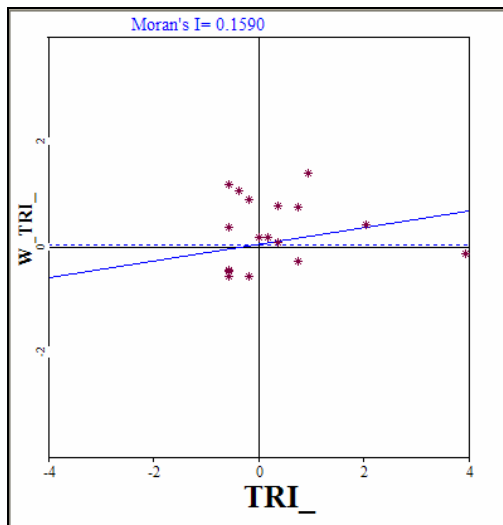


Figure 10. Moran's I value for Toxic substance releases. The graph shows the number of releases in a watershed (X) versus other study watershed values weighted by distance between watersheds (Y). The Moran's I = 0.1590 indicates no spatial auto correlation. Source: Chart developed using Geoda v 0.9.5 - i software.

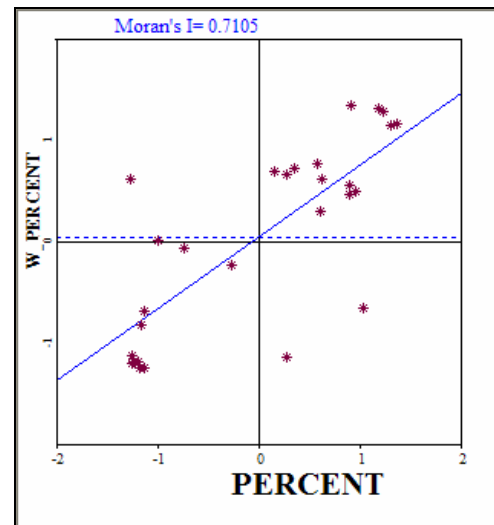


Figure 11. Moran's I value for non-point pollution sources. The graph shows average density of non-point pollution sources in a watershed (X) versus the average of other study watershed values weighted by distance (Y). The Moran's I = 0.71 indicates somewhat strong local spatial auto correlation. Source: Chart developed using Geoda v 0.9.5 - i software.

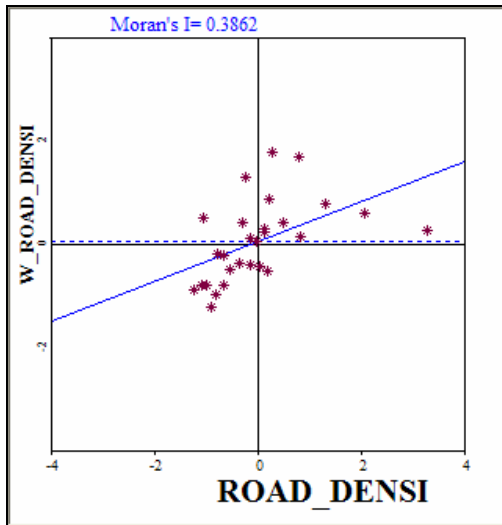


Figure 12. Moran's I value for Toxic road density. The graph shows the road density in a watershed (X) versus the average density in other study watershed values weighted by distance (Y). The Morans I = 0.47 indicates little local spatial auto correlation . Source: Chart developed Geoda v 0.9.5 - i software.

The following map is an example of clustering found in watersheds evaluated for non-point sources. This type of analysis would have been applied to all parameters givne more time.

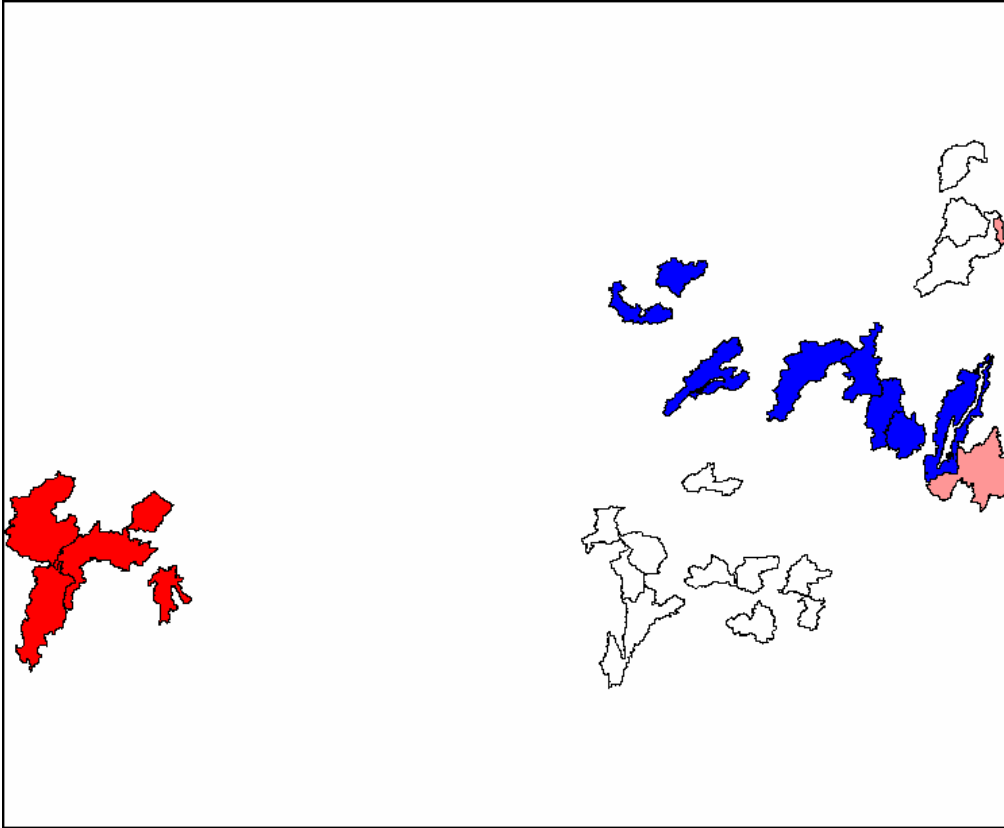


Figure 13. The map above illustrates the clustering of non-point data. Bright red illustrates a high individual percentage of non-point percentage in the watershed and in the nearest neighbors. The bright blue represents the opposite. The pink polygons are those watersheds that have high densities relative to the average of their 4 nearest neighbors. Source: Chart developed Geoda v 0.9.5 - i software.

No real surprises jump out from the first series of charts. The slope data may have been too insensitive to the range in slopes found across the state. Most of the watersheds indicated no slopes over 10%. Those that do occur are located mostly in the north east Adirondack Region. Topography does not appear to influence road density much, so there was not much spatial auto-correlation with respect to roads. The roads are sparse in agricultural areas as well as mountainous areas. Population density did show some autocorrelation, not surprising given that there are two large urban areas (one near Rochester, the other near Syracuse) and the rest of the area represented in the watersheds is rural. TRI sites appear to be focused in the Syracuse area, where some military bases and concentrated industry are located. The non-point values illustrate the high amount of agriculture in western NY (See the red in Figure 13 above as well). The Adirondack area experiences little agriculture activity (See the blue in Figure 13 above as well).

Conclusion

Some of the results of the analysis were expected, while some were surprising. As expected, a higher concentration of toxic release sites correlates strongly with impairment. The fact that this correlation is expressed even with the limited amount of watersheds represented in the analysis comes as not surprise as states focus large amounts

resources on detecting even the smallest traces of toxins. Any positive finding would trigger an impairment classification. This is due to the high risks to human health they pose. In addition to TRI, population also showed a relatively strong correlation with impairment class. This also would be expected. It was however surprising that slope was negatively correlated with impairment classification; that is, the more area in a watershed that contained slopes over 10%, the less likely it would be classified as impaired. Similarly, the magnitude of non-point sources and road density in the watersheds did not play more determinant role in impairment class.

Several factors could have affected the analysis outcomes with respect to slope, road and non-point stressors in the watersheds. First, impairment classification may have minimally or not sensitive to sediment. Additionally, the measurement of phosphates and nitrates in a fluvial environment requires the intensive sampling of storm events; the pollutants travel quickly through the stream system and may have been missed by the sampling. Specifically with respect to roads, most of them in New York are paved and therefore have rather stable surfaces and also they tend to have stable ditches. Therefore the sediment eroded directly from these features is probably minimal. That said, the increase of runoff from these structures may indeed lead to more water in nearby channels leading to stream channel instability and an increase in channel scour and deposition. Again, these effects would have no bearing on impairment classification if sediment were not assessed.

The ranking process used by some agencies to prioritize watersheds for improvement and protection needs further evaluation. One large drawback appears to be the ranking of the summation of individual parameter ranks. As impairment classifications are based on individual water quality parameters, it only one excessive value in one parameter to trigger a watershed being classified as impaired. In contrast, using ranking, a classification of impairment requires a high rank for numerous parameters, e.g. the ranking system depends upon summing the stressor values. The relationship determination would start with more closely delineating the watersheds that are impaired and stratifying them by impairment cause. Only then should the relationships between the stressor and rank be understood.

The current analysis indicates that a good bit of additional work is warranted if it is to lead to meaningful results. If asked to pursue this analysis further, I would expand the number of watersheds analyzed to include the entire state. I would then normalize the representation of impaired segments in streams to a density for each watershed. In this manner, I could then proceed to develop single and multiple regressions between the impairment density and the densities of the stressors.

In addition, complete watershed coverage would allow a better assessment of clustering and fragmentation. The analysis that was done illustrated some underlying first order spatial correlation, particularly in population and non-point. Many large populations were left out of the analysis (think New York City!). The inclusion would likely have led to a higher Moran's I for that variable. Also the on-point Moran's would also have probably been strengthened. In addition to the ones that did exhibit clustering in this

analysis, TRI may have started to exhibit higher values. Finally, a larger, more inclusive dataset would have allowed second order patterns, if they exist, to be found.

In the end, this analysis reinforces my closely held belief that there is nothing simple with regards to ecological/environmental relationships. GIS does offer a powerful tool to further explore extensive spatial and tabular datasets and a key to uncover solutions to the important ecological and environmental issues we face. In the end, this project has made it clear that environmental analysis is very complex.