Spatial distribution of land type in regression models of pollutant loading

Evan J. Fedorko[1], Robert Gilmore Pontius Jr.[1], Stephen P. Aldrich[2], Luc Claessens[3],
Charles Hopkinson Jr. [4] , and Wilfred M. Wollheim [5]

Abstract This paper proposes a method to improve landscape-pollution interaction regression models through the inclusion of a variable that describes the spatial distribution of a land type with respect to the pattern of runoff within a drainage catchment. The proposed index is used as an independent variable to enhance the strength, as quantified by R² values, of regression relationships between empirical observations of in-stream pollutant concentrations and land type by considering the spatial distribution of key land-type categories within the sample point's drainage area. We present an index that adds a new dimension of explanatory power when used in conjunction with a variable describing the proportion of the land type.

We demonstrate the usefulness of this index by exploring the relationship between nitrate (NO_3^-) and land type within 40 drainage sub-catchments in the Ipswich River watershed, Massachusetts. Nutrient loads associated with non-point source pollution paths are related to land type within the up-stream drainage catchments of sample sites. Past studies have focused on the quantity of particular land type within a sample point's drainage catchment. Quantifying the spatial distribution of key land-type categories in terms of location on a runoff surface can improve our understanding of the relationship between sampled NO_3^- concentrations and land type.

Regressions that employ the proportion of residential and agricultural land type within catchments provide a fair fit (R² = 0.67). However, we find that a regression adding a variable that indicates the spatial distribution of residential land improves the overall relationship between in-stream NO_3^- measurements and associated land types (R² = 0.712). We test the sensitivity of the results with respect to variations in the surface definition in order to determine the conditions under which the spatial index variable is useful.

Keywords

GIS, Non-point source pollution, nutrient export, spatial distribution, regression modeling

[1] Graduate School of Geography and George Perkins Marsh Institute, Department of International Development, Community and Environment, Clark University, Worcester, Massachusetts, EMAIL rpontius@clarku.edu, [2] Michigan State University, East Lansing, Michigan, [3] San Diego State University, San Diego, California, [4] Marine Biological Laboratory, Woods Hole, Massachusetts, [5] University of New Hampshire, Durham, New Hampshire

1 Introduction

1.1 Nitrogen Export

Nitrate, NO_3^-, is a major concern among hydrologists and water managers. In drinking water, elevated levels of nitrates can lead to serious illness and in some cases, death (EPA 2004). As part of the 1974 Safe Drinking Water Act, The United States Environmental Protection Agency set a maximum contaminant level goal for nitrates at 10 ppm. Health problems are not limited to human consumption, however. Nitrates in surface water contribute to the process of eutrohpication, which leads to high biological activity, in particular, algal growth. This, in turn, can choke other species, such as fish and small vertebrates (EPA 2004).

Most nitrates found in surface or groundwater have their origins in organic nitrogen or ammonia, generally from feces and urine that leak from septic systems or livestock feed lots. Organic nitrogen is converted to ammonium (NH_4^+) in soil through the process of mineralization. Ammonium, in turn, is converted to nitrite (NO_2^-) and nitrate (NO_3^-) by bacteria of the genus's *nitrosomonas* and *nitrobacter*, respectively (Pidwirny 2004). Inorganic potassium nitrate and ammonium nitrate are common ingredients in fertilizers, thus allowing the nutrient to skip any conversion process from organic nitrogen to ammonium to nitrate or nitrite (EPA 2003). Nitrate is very easily leached from soil and nitrogen deposition is very closely linked with hydrologic processes (Pidwirny 2004). Fertilizer application greatly accelerates the deposition of nitrogen and other nutrients in rivers, lakes and streams, and thus, accelerates eutrophication (Jordan 1997).

A myriad of studies relating heightened nutrient loads in surface water or groundwater to land type have been completed. Most commonly, studies of this sort concentrate on one of two land types, forest or agricultural land, depending on the region of study. The major cause of nitrate export into surface water and groundwater is inorganic fertilizer applied to crop land (Banasik 1999, Correll 1994, Jordan 1997, Mitchell 1999). The more intense the fertilization campaign, the greater the export of nutrients, nitrate in particular. Nitrate is leached from the soil and it enters the groundwater regime. Eventually, that nitrate appears in surface water.

Conversely, forest land is commonly thought to be a sink for NO_3^-. That is, given certain conditions, some nitrates can be removed from water that passes through these areas. Forested riparian areas are especially effective at removing nutrient loads from ground and surface water. Riparian areas are generally flat areas adjacent to stream channels. They are defined more concretely by three major characteristics: low slope, hydric soil type, and deep soils. Wetlands, an effective nutrient sink, are often found in riparian areas. Forested riparian areas and wetlands can be quite effective at removing nutrients from water, even during runoff events (Casey 2001). Forest land as a whole, inside or outside riparian areas, can effectively remove nutrients from water. Overall export of nitrogen from forested catchments is lower than from non-forested catchments (Band 2001). Wollheim et al. (2004) determined that increases impermeability

increases nitrogen loading through three factors: an increase in N sources, increased base flow, and a decrease of contact between runoff and soils.

Scientists have used regression to quantify relationships between land type and nutrient export to predict future nitrate concentrations. Combining these relationships with predictions of future change in land type will enable researchers to make informed speculations about future water quality. Within the Ipswich River Watershed, Pontius et al. (2000) extrapolated a land-type change scenario to 2101 and used an empirical nitrate loading/land type relationship to predict that nitrate loading would increase as land was converted from forest to residential land. Their study used a distance weighted regression to determine nitrate loading potential per land type. While still a developing science, land-type change modeling may benefit from the use of spatial statistics when fleshing out impacts based on predictions.

1.2 The Ipswich river watershed

Nitrogen export is a primary environmental concern in the Ipswich River watershed. The Ipswich River drains into a large, estuary and protected saltwater sound – the Plum Island Sound. The Ipswich River watershed is located in the coastal lowland section of New England, in northeastern Massachusetts (Figure 1). In the early 1600s, Captain John Smith described the Ipswich River as a "land of promise" when he first laid eyes on the river valley. People apparently agreed with him. Today, there are 21 towns that touch or are within the watershed's boundary. Within those towns, the river serves as a water source to 168,000 people and thousands of businesses (Census 2000). Use is not limited to towns within or touching the watershed, however. In total, 330,000 people depend on the river for drinking water (Bowling 2003). Permitted water withdrawals averaged 114.7 million liters per day in 1999. American Rivers, a non-profit river conservation organization, placed the Ipswich River at number 3 on their 10 Most Endangered Rivers list of 2003 (American Rivers 2003, Bowling 2003).
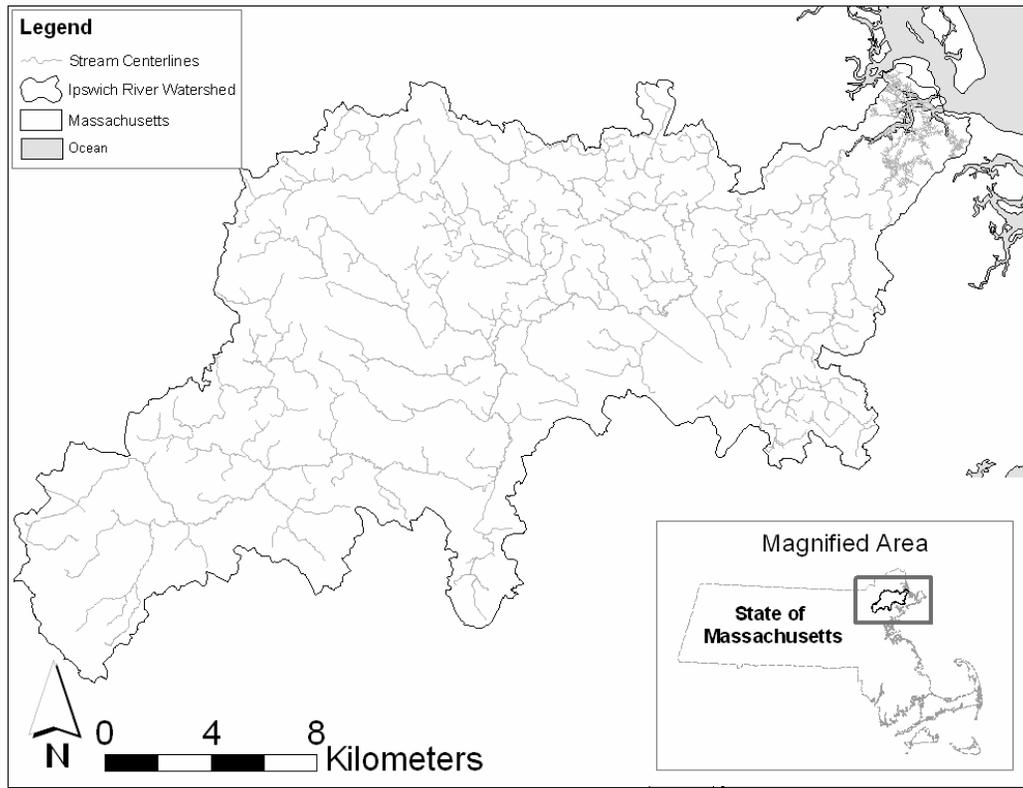
Figure 1. Ipswich river watershed within Massachusetts, USA.

The peak elevation in the watershed is 125 meters. The Ipswich River watershed falls only 75 meters from the headwaters to the Ipswich Dam, then another 40 meters in the last few miles. The slope is as steep as 36°, but remains relatively flat, with an average slope of only 3.5°. The bedrock lithology of the watershed is largely mafic or granitic with some metamorphic rocks. Roughly half (49%) of the surface geology in the watershed is glacial till or bedrock. The remaining portion is mostly sand and gravel (42%) with small portions composed of flood plain alluvium (7%) or fine grain deposits (2%). Overlying this are mostly shallow, poorly drained soils (Baker 1964, Wollheim 2004). Landuse data from 1999 tells us that the watershed is largely forested (35%) with residential land and wetland making up the lion's share of the remaining area with 30% and 20%, respectively.

The watershed is under increasing development pressure. The Massachusetts Department of Environmental Protection observed a population increase in the watershed of 9% between 1980 and 2000, while the proportion of residential land within the watershed increased by 35%. This has come at the expense of forest and forested wetlands within the watershed, decreasing those land-cover types by 15% and 25%, respectively (Bowling 2003). Our observations of land type change between 1985 and 1999 (Table 1), while less drastic, reflect similar overall trends. This increased population has led to increasing demand for water. Low flow

conditions in the watershed have become more frequent. As much as half of the main stem of the river ran dry during the summer months of 1995, 1997, 1999 and 2002 (Bowling 2003). Low flow conditions have exacerbated problems in the watershed and led to heightened attention on the river as a resource.

Table 1. Proportional land type changes in the Ipswich River watershed between 1985 and 1999. Values given are percentage of total area.

Land Type	1985	1999	Change
Agriculture	8.91	8.03	-0.88
Forest	52.75	49.00	-3.75
Wetland	5.01	4.71	-0.31
Industrial-Commercial	3.40	3.77	0.37
Residential	26.52	31.10	4.58
Salt Marsh	1.23	1.23	0.00
Water	2.17	2.16	-0.01

1.3 The importance of spatial analysis

Many traditional methods used to explore relationships between land type and non-point source water pollution have relied on one simple variable i.e., proportion of land type within a drainage catchment. While this type of analysis has the advantage of being easy to measure and interpret, it ignores a potentially very important consideration, the spatial arrangement of the land-type within the catchment. Intuitive knowledge of flow regime and nutrient export tells us that flow through and near residential land will result in a very different nutrient export than flow farther from residential land.

Spatial arrangement of land types can be important to understand nutrient sinks also. Forested riparian areas have the potential to act as denitrification zones, even in urbanized areas (Groffman 2003). Riparian forests that are located such that they receive a large amount of runoff from residential areas should remove more nitrogen from the hydrologic system than those concentrated in low drainage areas far and upstream from nutrient sources. Therefore, any type of analysis that depends solely on a variable measuring the proportion of these or other land types would be unable to recognize that forested riparian areas are very different than forested hillsides in the context of pollutant loading. This failure could be important because forested riparian areas have the ability to remove comparatively larger amounts of nitrate from surface and groundwater through denitrification (Groffman 2003). The denitrifying ability of any land type is limited by the potential for N-saturation (Band 2001, Wollheim 2004). Natural systems and processes are not random, but rather, demonstrate spatial structure or arrangement (Dale 2002). Riparian areas are defined in spatial terms. We propose that spatial structure is inherently linked to variation in non-point source nutrient loading in surface water and we provide an immediately accessible method of measuring spatial arrangement of land types for inclusion in analyses.

Spatial analysis is commonly defined as "a general ability to manipulate spatial data into different forms and extract additional meaning" (Bailey 1994). It includes a broad array of tools that range from relatively simple measures of distance and proximity, such as nearest neighbor analysis, to more complex statistics such as the Moran's I, a measure of dependence across space, often referred to as spatial autocorrelation (Shekhar 2003, Isaaks 1989). For the purpose of this paper, we focus spatial analysis on statistical methods, as discussed by Bailey (1994), who limits the discussion of spatial analysis to methods that take into account the stochastic nature of patterns. Statistical spatial analysis seeks to quantify spatial patterns on a landscape. In this paper, we present a descriptive index that quantifies spatial distribution of a binary phenomenon (i.e. a single category of land type) across many catchments. We design the index so that it is useful for a variety of applications while we illustrate its usefulness by applying it in a regression analysis for the Ipswich River watershed. Potential applications of this index are wide and cover a variety of surface and landscape processes.

2 Methods

2.1 Data

There are three primary types of information in this analysis: in stream samples of nitrate, digitally mapped categorical land-type data, and GIS-generated surfaces. The nitrate data was collected in February of 2000 as part of a sampling campaign undertaken by scientists of the Marine Biological Laboratory in Woods Hole, Massachusetts. Samples were analyzed for nitrate on a Lachat autoanalyzer using a cadmium reduction column.

GPS points of the sampling points were collected as part of the campaign. These sample points were used to generate sub-catchments (Figure 2). We supervised the creation of reliable sub-catchments for the headwater non-nested sample points. These sub-catchments were used as study extents to compute the proportional amounts of land-type categories in each sub-catchment. The land-type data was assembled by aggregating data of 37 categories supplied by the Massachusetts Executive Office of Environmental Affair's Massachusetts Geographic Information System into seven categories. A wetlands dataset, created by the University of Massachusetts/Amherst and field checked by the Massachusetts Department of Environmental Protection as part of the Wetlands Conservancy Program, was then combined with the seven category data to create a final, wetland corrected land-type dataset for 1999. The five categories of land-type that we consider in our analysis are: forest, wetlands, industrial/commercial, residential and agriculture. We computed the proportions of each land type within each catchment.
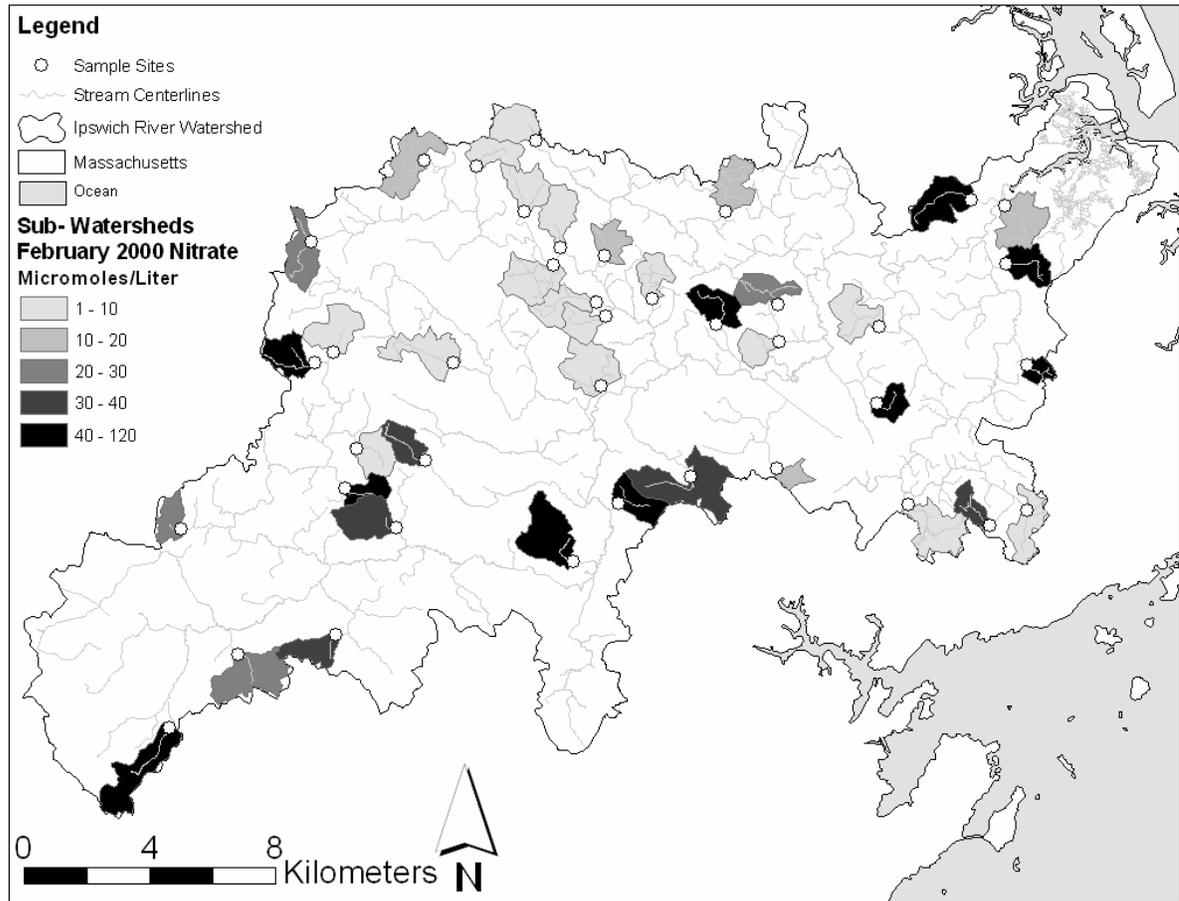
Figure 2. Forty headwater catchments within the Ipswich river watershed.

We also calculated a variable describing the density of septic systems with each sub-catchment. This variable was derived from Census 2000 data. Each census block contains an attribute describing the proportion of households within the block that use septic systems. We calculate the number of households per unit area within each sub-catchment that have septic systems and include it in our analysis.

We generated three surfaces for this analysis: a pair of surfaces, raw runoff and logarithmically transformed runoff, based on a 30 meter digital elevation model (DEM), and an additional surface based on Euclidean distance. The DEM was provided by the Massachusetts Geographic Information System (Mass. GIS 2005). The original five meter DEMs are derived from photogrammetric points and breaklines collected from 1:5000 scale orthophotgraphy. The provided elevation surface was resampled from its original resolution of five meter pixels to 30 meter pixels. Before creating the analysis surfaces, we enhanced the DEM with hydrologic data. First, we created a "depressionless" or "pit-free" DEM with Arc Map's "fill sinks" tool. Following this, we accounted for hydrologic channels by burning 1:100,000 scale stream centerlines into the

DEM. This was accomplished by subtracting a raster map of the stream centerlines from the DEM, which produced channels in the DEM's surface. This allows for the calculation of a hydrologically adjusted routing of runoff using ARCINFO GRID.

The runoff surface was created in two steps. First, a flow direction surface was created using the hydrologically corrected DEM as an input. In this surface, each pixel has one of 8 possible values, each corresponding to a direction of flow into an adjacent cell. The value indicates the direction that water would flow upon exiting that cell. Next, we created a runoff surface using the flow direction raster as an input. Each cell in the runoff surface has a value that indicates the number of cells that flow into that cell. As the number increases, so does the predicted amount of water that passes through that cell. This constitutes the first raw runoff map. We created a second runoff surface by taking the natural logarithm of the raw runoff map. On both of these runoff surface maps, cells that are nearer the sample site or drainage channel have larger values, where nearer is defined with respect to runoff.

One of the primary aims of this research is to present a simple, but statistically sound method of accounting for spatial distribution of features across a surface. To that aim, we included the logarithmic transformation of the raw runoff surface in our analysis in order to adhere to the conventions of statistical analysis. Logarithmic transformation is a commonly used method in statistical analysis as it aids in the satisfaction of the assumptions of symmetry and homoscedasticity while improving linearity. These are all desirable conditions and often aid in the performance of statistical techniques. We were very interested in how the spatial index might be improved using conventional statistical transformations.

The third surface is the Euclidean distance from each of the 40 nitrate sample points. In this surface, values near the sample site have smaller values and near is a function of straight line distance.

2.2 Spatial Index

2.2.1 Definition

The spatial distribution index, Equation (1), is a number that describes the distribution of a land type (or any other binary geographic characteristic) across a surface within a specified geographical extent. In this case, the surface is one of the three surfaces described in the previous section and the geographical extents are the 40 sub-catchments derived from surface-water sample points. Equation (1) returns a descriptive value for each sub-catchment.

$$SI = \frac{\sum_{n=1}^{N} W_n S_n P_n}{\sum_{n=1}^{N} W_n S_n \overline{P}}$$ **Equation (1)**

$$\overline{P} = \frac{\displaystyle\sum_{n=1}^{N} W_n P_n}{\displaystyle\sum_{n=1}^{N} W_n} \qquad \textbf{Equation (2)}$$

Equation (1) defines SI as the spatial index for the catchment. Lower case n is the index for each cell in the catchment, while upper case N is the number of pixels in the catchment. W_n is the weight (between 0 and 1) for pixel n. For our example, W_n = 1 for all n, but the weight is included as a variable to make the equations general. S_n is the value on the selected surface map for cell n and P_n is the proportion of a land type in pixel n. $\overline{P}$ is the proportion of the land type in the catchment, given by Equation (2).

If the land type is distributed evenly within the catchment, then the value of I is 1. If the land-type is concentrated in areas of low values on the surface, e.g. low drainage far from the sample site or drainage channel, then the value of I is less than 1. If the land type is located primarily in areas of high surface values, e.g. high drainage near the sample site or drainage channel, then I is greater than 1 (Aldrich 2002).

One feature of I is that it is statistically and conceptually independent from $\overline{P}$ (Aldrich 2002). This is desirable because a regression analysis with proportion of a land type as an independent variable will be unlikely to encounter problems with multi-collinearity when it uses the spatial drainage index of the land type as an additional independent variable. This characteristic is illustrated by near zero measured association (R^2 = 0.000) between proportion residential and the residential spatial drainage index for the 40 catchments in Figure 2.

2.2.2 Examples

The spatial index seeks to illuminate the difference between the catchments shown in Figures 3 and 4, called catchments A and B, respectively. Forty-three percent of the variation in February, 2000 NO_3^- measured in the Ipswich River watershed can be explained by residential type land-cover (Figure 5). Residential land constitutes 35% of the area within both the sub-catchments pictured in Figures 3 and 4. However, NO_3^- measured in February of 2000 was 20 µm/l in catchment A, while NO_3^- was nearly three times higher, 73µm/l, in catchment B. Initial visual inspection indicates that the arrangement of residential land within the catchments is different. It appears that catchment A has more residential land within close proximity of both the sampling point and the stream channel than catchment B, which seems to be buffered from residential land along the stream channel. This potentially biased observation contradicts the hypothesis that higher nitrate would be observed at sample sites that are near residential areas. The spatial index (I) is designed to quantify the arrangement of land type with respect to each of the three surfaces: the raw runoff surface, the log-transformed runoff surface, and the Euclidean

distance from the sample point surface. It is important to calculate the spatial index because the human eye can be fooled, as it is difficult to consider simultaneously the land type map as well as the surface map.
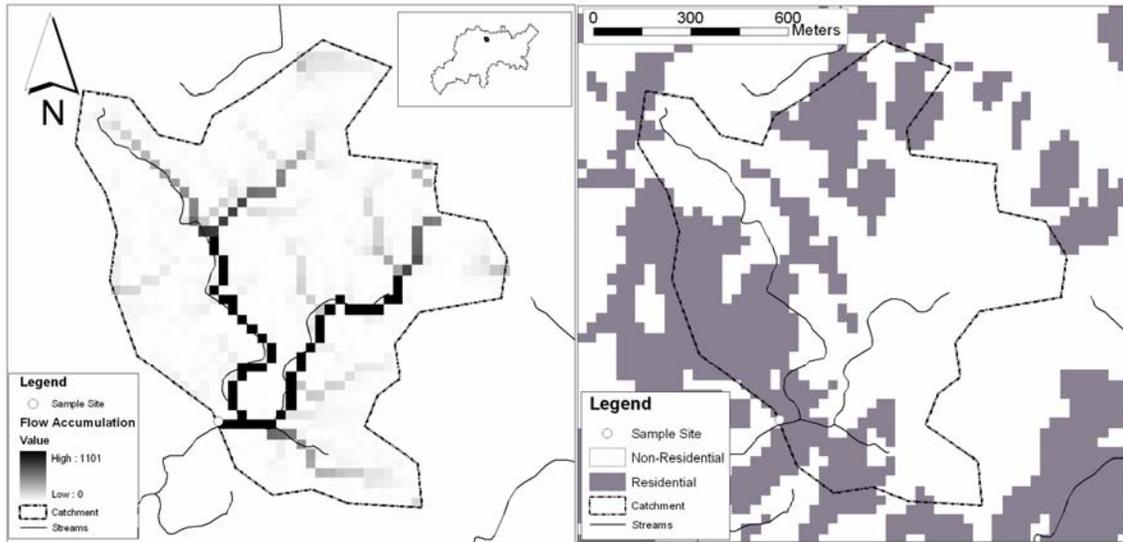

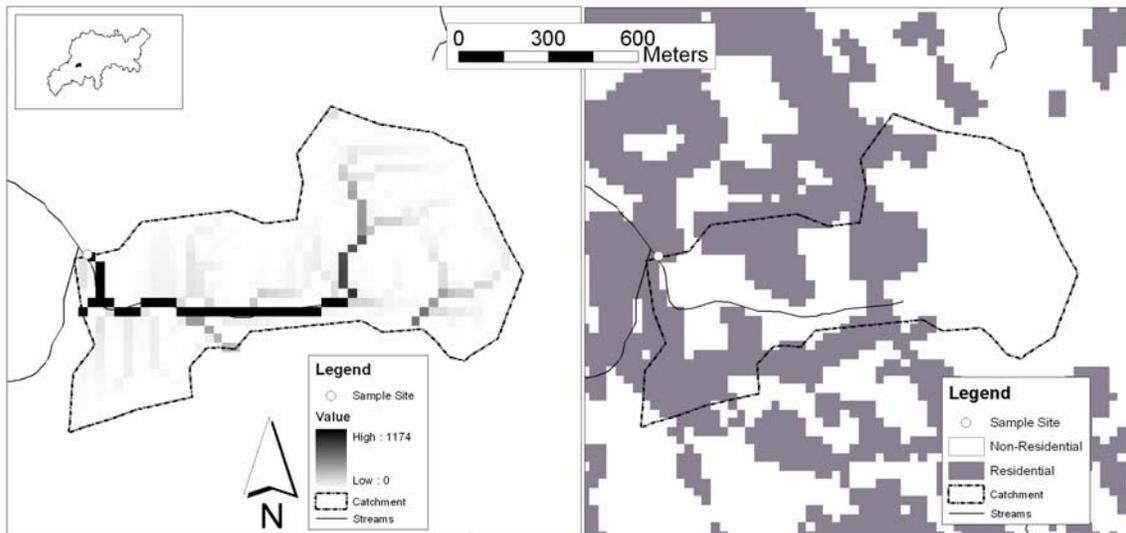Figure 3. Catchment A, with a low raw spatial drainage index of 0.17.

Figure 4. Catchment B, with an even raw spatial drainage index of 1.03.

With respect to the raw runoff surface, catchment A has a very low spatial drainage index (0.17). This tells us that residential land is concentrated in areas of low drainage. Conversely, catchment B (Figure 4) has a spatial drainage index of 1.03, indicating that residential land is more evenly distributed, slightly concentrated in areas of higher drainage. This fits our hypothesis that higher values of the spatial drainage index will align with higher NO_3^- values.

We can see some of the differences between the catchments that are being quantified by the spatial index. Starting with Figure 3, note the location of gray pixels on the right hand portion of Figure 3. These are pixels classified as residential land. In particular, note the large block running from the northwest to the south along the edge of the sub-catchment near the pour point. A naïve analysis would suggest to us that, since so many residential pixels are close to the pour point, that the value of this index will be very high. However, if we compare this large block to the runoff map, on the left side of Figure 3, we can see that much of this large block overlays very light, almost white pixels, indicating low runoff values. Even the portion immediately southeast of the pour point is in an area of very low drainage. Examining the rest of the map, we can see that only a small portion of the residential pixels touch high drainage pixels on the runoff map.

Compare this with Figure 4. At first glance, it would appear that, like catchment A, very few residential pixels intersect the stream channel. However, examining the runoff map, we see a large runoff channel not evident in the vector streams that extends almost due north from the vector stream channel. This area of high drainage is also an area of residential land. Likewise, the narrow mouth of the catchment contains a great deal of residential land. Other parts of the residential land, e.g. the southwest corner, are clearly in areas of low drainage. Residential land is distributed such that about half of it lies in areas of higher drainage, and half in lower drainage. Hence, we find a spatial drainage index near 1, i.e. 1.03.

2.2.3 Sensitivity Analysis

The spatial index is sensitive to the definition of the surface. That is, one catchment will return different index values for the same land-type depending on which surface is used to calculate the index. These different values will, in turn, result in different results in statistical analysis.

In order to explore how different surface maps affect the outcomes of regression models, we create three indexes for each drainage catchment and use them in regressions. We calculate the index using three surfaces: a raw runoff surface, a logarithmic transformation of the raw runoff surface, and a Euclidean distance from the sample point surface. Hence, we calculate two spatial drainage indices and one spatial distance index. The indices are to be interpreted in different ways. If we use a runoff surface, a high value of the spatial index signals that the land type is concentrated in areas of high drainage, while a low index value signals that the distribution is in areas of low drainage. If we use the index with a Euclidean distance map, a low value of the spatial index indicates that the land type is concentrated in areas close to the drainage point, while a high index value signals that the distribution is far from the drainage point. Each

measurement of land-type proportion is paired with each of its three spatial indexes to assess the influence on the regression.

2.3 Regression

Stepwise regression considered the proportion and spatial indexes of the following: residential, forest, wetland, industrial/commercial, and agricultural. The analysis also included the mean septic system density of each catchment. We used the stepwise method of entering variables with a 0.05 p-value tolerance for inclusion of the variable in the regression. We ran a total of four stepwise regressions. First, we used only the land-type proportion variables and septic system density to determine the best non-spatial regression. Next, we used all the variables previously stated, as well as the spatial drainage indices based on a raw runoff surface for each of those land types. The next two stepwise regressions each included one spatial index type in place of the raw runoff spatial index, first we included those calculated with the log-transformed runoff map and then those calculated based on Euclidean distance surfaces. In order to complete the sensitivity analysis, regressions were constructed that paired each proportional variable with each of its three spatial indexes to examine changes in R² values as the surface definition changes.

3 Results

Regression results show that the best single explanatory variable is percentage residential (Figure 5). Consideration of the spatial arrangement of land categories gives modest increase in goodness of fit. Table 2 gives R² values from stepwise regressions in the right most column and coefficients with p-values of included variables in the middle columns for that analysis that uses the raw runoff surface. Two key index land types are agriculture and residential. Regressions (b) and (c) show that proportion agriculture explains 22% of the variation in NO_3^- and residential explains 43%. In both of these regressions the variables are significant at the 0.02 level. Conversely, regression (a), using only the spatial drainage index for residential land produces a low R² value with a p-value of 0.407.
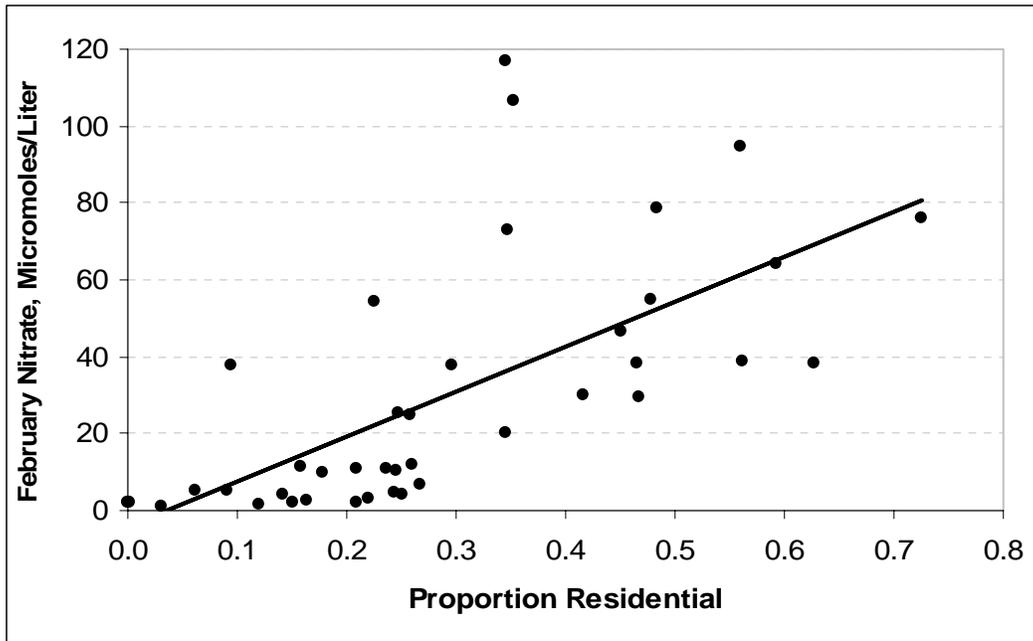
Figure 5. Nitrate versus proportion residential land-type by catchment with least squares line. R²
= 0.434 with a p-value of 0.000.

In regression (e), proportion residential is paired with the residential spatial drainage
index. This produces an R² value of 0.469, an improvement of over proportion residential by itself.
The p-value of the residential spatial drainage index is not significant (p =0.214). Regression (f) is
the best non-spatial regression. This regression uses proportion agriculture and proportion
residential as independent variables. Both variables have 0.000 p-values. These two variables
explain 67% of the variation in the dependent variable. Regression (g) increases the R² to 0.712
by adding the spatial index. The p-values of all three variables are significant at the 0.03 level.
Including a measurement of the spatial distribution of residential land results in an improvement in
the goodness of fit.

Table 2. Regression coeffecients, p-values and R² values.

Run	Percent Agriculture	Percent Residential	Residential Raw Spatial Drainage Index	R²
a.	n.i.	n.i.	14 (0.407)	0.02
b.	187 (0.002)	n.i.	n.i.	0.22
c.	n.i.	117 (0.000)	n.i.	0.43
d.	190 (0.001)	n.i.	17 (0.273)	0.25
e.	n.i.	119 (0.000)	20 (0.214)	0.47
f.	119 (0.000)	194 (0.000)	n.i.	0.67
g.	122 (0.000)	199 (0.000)	23 (0.028)	0.71

Figure 6 illustrates how R² values change as a function of which surface map is used to calculate the spatial index. For each of three land-types, only the proportion variable is used initially to explain variation in February 2000 NO_3^-. Following that, each proportion is paired with a version of its respective spatial index. Therefore, each land-type has a total of four regressions. R² values are sensitive to which surface the spatial index uses. In the case of forest, only one surface map, Euclidean distance from the sample point, adds any explanatory power. The other two land types exhibit similar behavior in that they exhibit sensitivity to the selection of the surface map.

4 Discussion

4.1 Interpretation

The variation in regression results, illustrated by Figure 6, shows that it is important for a scientist to understand the underlying surface being used to calculate the spatial index. Likewise, it is important to examine critically the relationships that are revealed from different indexes. The forest land type demonstrates a large increase in R² values when paired with a distance index, but demonstrates no relationship with drainage indexes. The wetland land type demonstrates a similar behavior in that the best pairing is with a distance index.
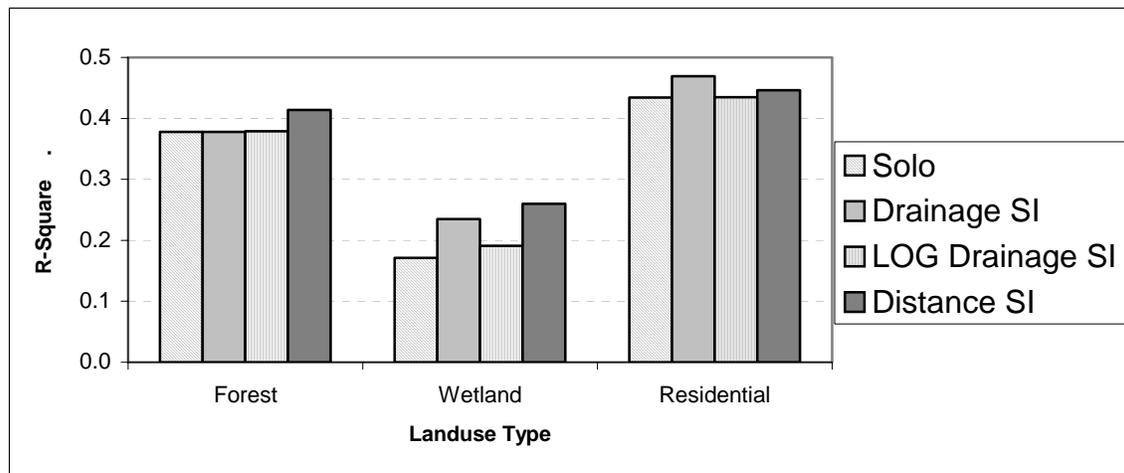
Figure 6. R² values as a function of spatial index definition for regressions with February 2000 NO_3^- as the dependent variable.

Figure 7 shows how the indices vary as a function of the surface map used to calculate the index for the same set of 40 catchments. Most catchment's index values in all three plots are scattered above the one to one line. Figure 7(a) reveals a slight positive relationship between raw runoff index values and log runoff index values (R² = 0.172). Overall, a logarithmically transformed drainage map produces values that are closer to 1 than does an untransformed drainage map. On the raw drainage map, cells close to the pour point and along the stream channel have a large amount of influence on the index due to their comparatively large values. The log transformation results in a more equitable distribution of influence among the cells within the catchment, thus pushing index values closer to one and reducing the variation among the catchments. Catchments with raw runoff index values that are less than 1 tend to have larger log drainage index values. Catchments with raw runoff index values greater than 1 tend to have smaller log drainage index values. Transforming the raw runoff map logarithmically results in a decreased range of index values compared to values derived from the raw runoff map. Knowledge of only raw runoff values allows us to make some generalizations about index values calculated with a log runoff map. However, log runoff index values are unique enough that they provide additional information about the nature of drainage landscape.
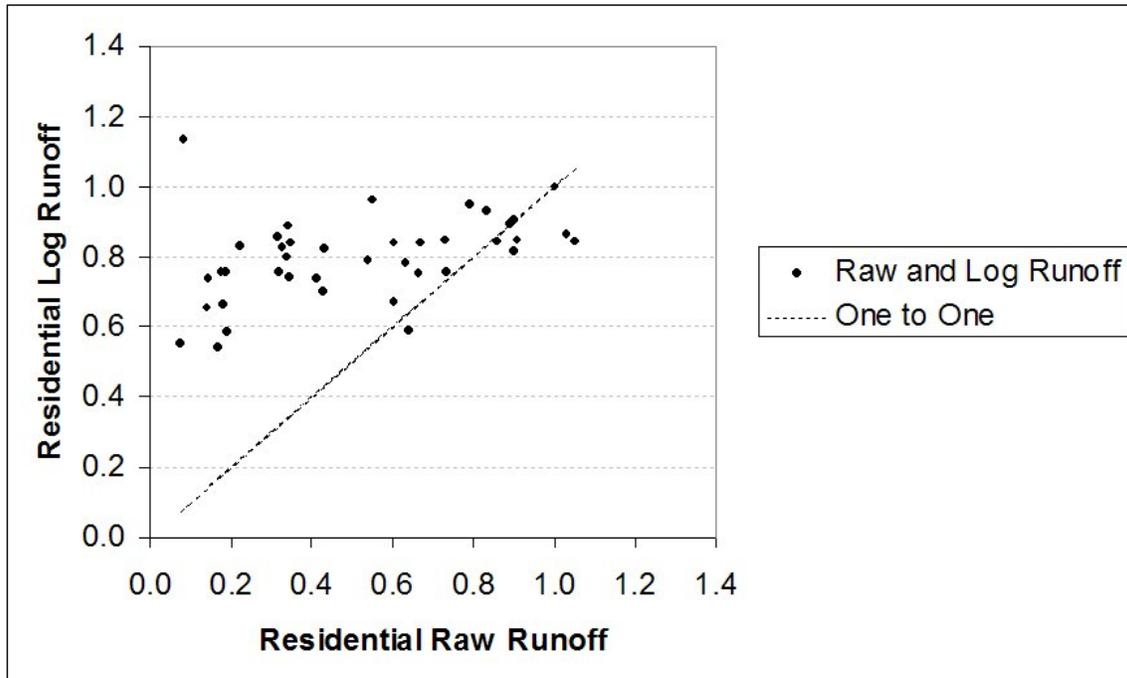
Figure 7 (a). Comparison of raw runoff and log runoff spatial index values.

Figures 7(b) and 7(c) compare each of the two runoff indexes with the distance index. Unlike what we observed in Figure 7(a), we can see no linear relationship exists between the runoff indexes and the distance index. This tells us that the distance map and resultant index values are telling us something fundamentally independent about the distribution of land type with each catchment. The distance index values tend to be larger than those produced with runoff maps. This reflects reality. Distance index values are evenly spread around a value of 1 while raw runoff values tend to be less than 1. This makes sense because, for legal and practical reasons, people do not live in or close to stream channels.
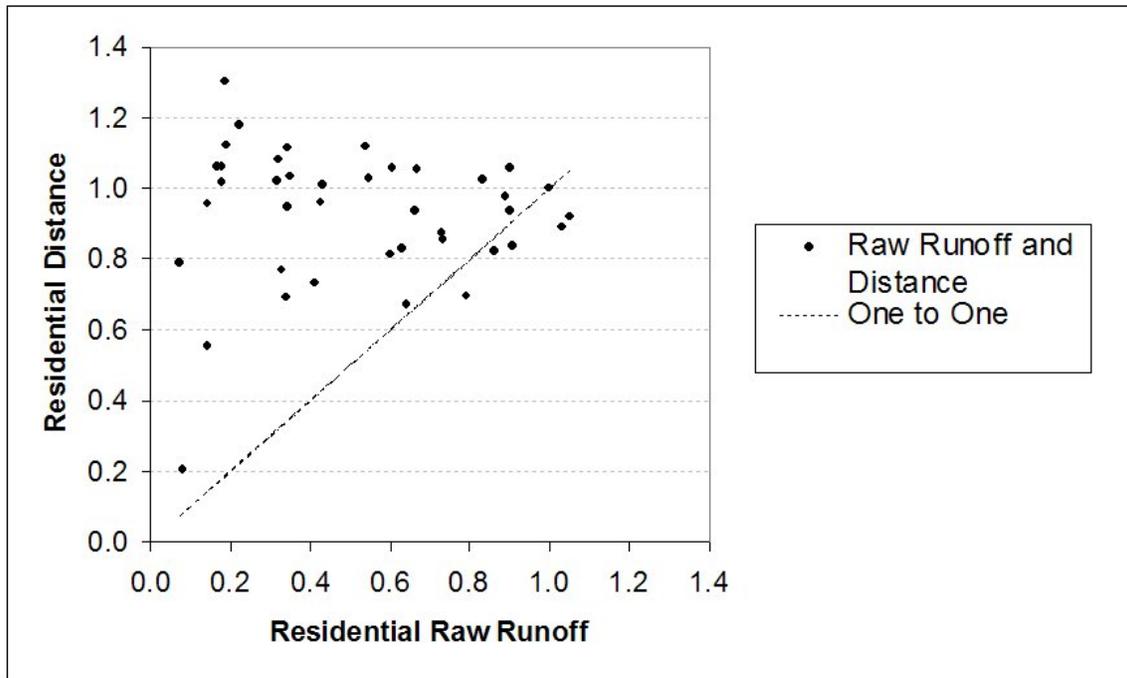
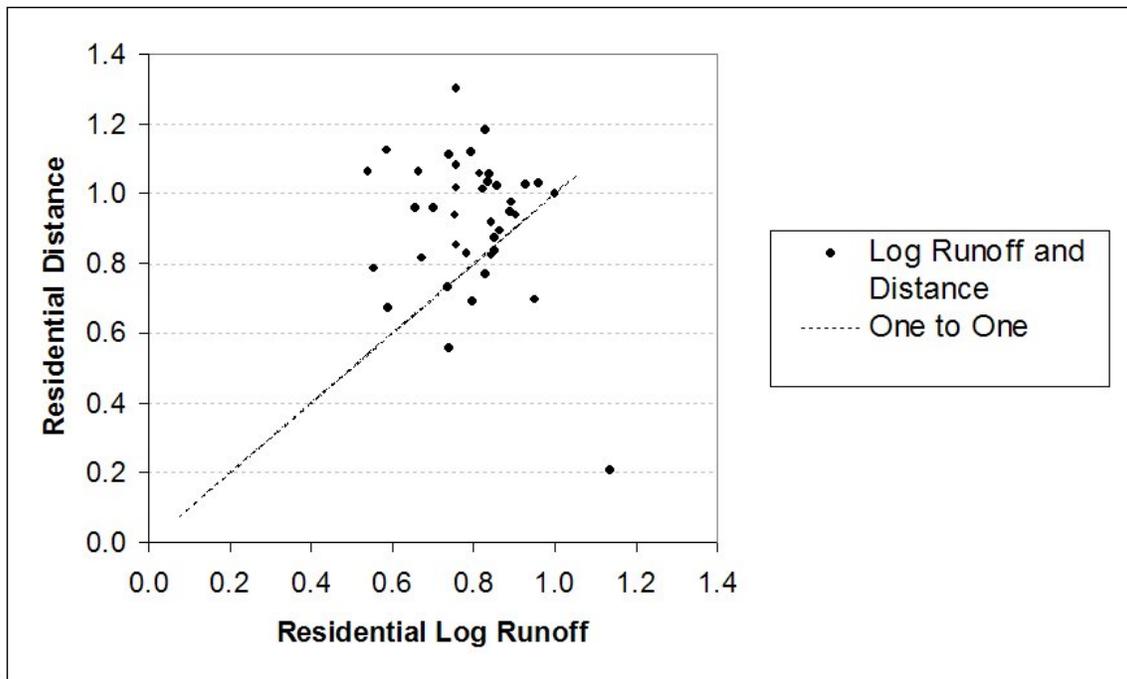Figure 7 (b). Comparison of raw runoff and euclidean distance spatial index values.

Figure 7 (c). Comparison of log runoff and euclidean distance spatial index values.

A scientist interested in using spatial index values in a regression analysis should use at least two of the index calculation methods presented here. One regression should use index values from a Euclidean distance surface another regression should employ index values from a runoff surface: raw, logarithmically transformed, or both. Distance and drainage maps provide independent information. Log transformation should dampen the effect of outliers that may exist in a raw runoff map.

All three of the variables in regression (g) of Table 2 have a positive relationship with NO_3^- measured in surface water. Agriculture is a well-documented source of nitrate, and residential land is becoming a new and potent source, due in large part to suburban lawn fertilization (Wollheim et al. 2004). It is interesting that the spatial drainage index for agriculture was not included in the final stepwise regression. This is probably due to a few characteristics of agricultural land. First and foremost, only 7% of the land in the watershed is agriculture and, on average, only 5% of each sample catchment is agriculture. Most catchments have very low values for agriculture's spatial drainage index. Thus, the location of agriculture does not indicate anything additional about agriculture's relationship with nitrate, because there is not much variation in the independent variable.

Multicollinearity is one of the dangers of regression models that consider proportion of more than one land types. The proportions within a sub-catchment must sum to one, and therefore they must demonstrate collinearity. Additionally, regressions that use both forest and residential land types will result in some collinearity because these two categories dominate the Ipswich River watershed. We tested for collinearity and found that it is not severe. The variance inflation factors are all very close to 1, indicating that no substantial collinearity exists.

4.2 Next steps

This study illuminates relationships between land type and nitrate discharge from headwater catchments in the Ipswich River. The Ipswich River watershed has been the subject of several land type change studies. A next logical step in the progression of research would be to integrate the coefficients produced in this study into a forecasting exercise that takes into account potential land-type change as it relates to potential changes in nutrient discharge.

The spatial index can be used to gauge distribution of a binary phenomenon across any surface, whether it is a runoff surface, a Euclidean distance surface, or some other surface. Jordan et al. (1997) found strong indicators that nitrate export is closely linked with groundwater flow. This supports the hypothesis that nitrate export is related directly to the amount of impervious surfaces in a drainage catchment (Wollheim et al. 2004). This is due to the fact that impervious surfaces have two effects: first, impervious surfaces increase runoff, and second, they reduce the amount of runoff that "touches" soil where nitrogen processing would take place.

Their two studies suggest that an examination of the spatial distribution of certain land types, i.e., residential and agricultural, across a permeability surface might reveal an important

relationship. According to Jordan et al. (1997), residential land in highly pervious soils which are most susceptible to leaching, will export more nitrate to groundwater, and hence, surface water. Almost conversely, Wollheim's hypothesis suggests that imperviousness will increase nitrate export, so perhaps residential land-type concentrated close to impervious surfaces will become a strong source. Our spatial index can be used to resolve these conflicting hypotheses by increasing the sophistication of empirical analysis.

A very appropriate application of the spatial index in the context of nutrient loading would be to design landscapes in which land types that effectively remove nutrients from water are placed in high drainage areas to minimize nutrient export. Sharma et al. (2001) performed a study in which maps were altered digitally in order to increase water retention. Similarly, scientists may be able to inform zoning practices by treating conceptual landscapes and examining those landscapes in the context of their distribution across runoff surfaces, such that sustainable solutions to nutrient loading problems are envisioned.

5 Conclusions

We have presented a methodology for calculating an index that measures the spatial distribution of a binary characteristic, such as a land type, with respect to any surface. We have provided examples of surfaces and distributions that this statistic can calculate. We have demonstrated that this index can add a degree of explanatory power in statistical analysis of the interaction between land type and non-point-source pollutants when paired with variables describing the quantity of the land type being measured. The methods presented in this paper can be used by scientists, researchers and planners to easily quantify the spatial distribution of features on the landscape and to explore how those distributions affect a variety of landscape processes.

Acknowledgements

Literature

Aldrich, J.H. and Nelson, F.D. (1984) Linear probability, logit, and probit models. Newbury Park, CA: Sage Publications.

Aldrich, S.P. (2002) Water quality as a function of the spatial distribution of land use. Honors Thesis. Worcester, MA: Clark University.

American Rivers (2003) America's most endangered rivers of 2003. Washington DC: American Rivers.

Bailey, T.C. (1998) Review of statistical spatial analysis in GIS. In Spatial Analysis and GIS, edited by S. Fotheringham and P. Rogerson, pp.13-44. Philadelphia: Taylor and Francis.

Baker, J., Healy, H., Hackett, O. M. (1964) Geology and Ground-Water Conditions in the Wilmington-Reading Area of Massachusetts 1694. Washington D.C.: USGS.

Banasik, K., Mitchell, J.K., Walker, S. E., Rudzka, E. (1999) Comparison of nutrient outputs from two lowland watersheds with different agricultural practices. In: Impact of land-type change on nutrient loads from diffuse sources, edited by L. Heathwaite, pp. 13-16. Oxfordshire, UK: IAHS

Band, L. E., Tague, C.L., Groffman, P., Belt, K. (2001) Forest ecosystem processes at the watershed scale: hydrological and ecological controls of nitrogen export, Hydrological Processes, 15, 2013 - 2028

Bicknell, B. R., Imhoff, J.C., Kittle Jr., J.L., Donigian, Jr., A.S., Johanson, R. C. (1997) Hydrological simulation program--fortran, user's manual for version 11, EPA/600/R-97/080. Athens, GA.: Ecosystems research division.

Bowling, A., Mackin, K. (2003) State of the Ipswich river, 2003. Topsfield, MA: Ipswich river watershed association.

Casey, R.E., Klaine, S.J. (2001) Nutrient attenuation by a riparian wetland during natural and artificial runoff events, Journal of Environmental Quality, 30, 1720-1731.

Correll, D. L., Jordan, T. E., Weller, D. E. (1994) The Chesapeake bay watershed: effects of land-type and geology on dissolved nitrogen concentrations. Conference Proceedings, Toward a sustainable coastal watershed: the Chesapeake experiment. pp. 639-648.

Dale, M.R.T., Dixon, P., Fortin, M., Legendre, P., Myers, D.E., Rosenberg, M.S. (2002) Conceptual and mathematical relationships among methods for spatial analysis, Ecography, 25, 558-577.

Environmental Protection Agency (2003) Eutrophication, Mid-Atlantic Integrated Assessment, http://www.epa.gov/maia/html/eutroph.html (July 2004).

Environmental Protection Agency (2004) Consumer fact-sheet on nitrates/nitrites, Groundwater and drinking water, http://www.epa.gov/safewater/contaminants/dw_contamfs/nitrates.html (July 2004)

Fedorko, E.J., Pontius Jr., R.G., Aldrich, S.P., Claessens, L., Hopkinson Jr., C., Wollheim, W.M. (2004) Spatial distribution of land type in regression models of pollutant loading, abstract. Biological Bulletin, 207.

Groffman, P. M. and Crawford, M. K. (2003) Denitrification potential in urban riparian zones, Journal of Environmental Quality, 32,1144-1149.

Isaaks, E.H., Srivastava, R.H. (1989) An introduction to applied geostatistics. Oxford: Oxford University Press.

Jordan, T.E., Correl, D.L., Weller, D.E., (1997) Relating nutrient discharges from watersheds to land-type and streamflow variability, Water Resources Research, 33, 11, 2579-2590.

Massachusetts Geographic Information System (2005) Digital Elevation Model (1:5000), http://www.mass.gov/mgis/img_elev5k.htm (2005)

Mitchell, J.K., Walker, S.E., Hirschi, M.C., Cooke, R.A.C.(1999) Nitrate losses under various cropping systems. In: Impact of land-type change on nutrient loads from diffuse sources, edited by L. Heathwaite, pp. 33-39 Oxfordshire, UK: IAHS.

Pidwirny, M. (2004) The Nitrogen Cycle, Fundamentals of Physical Geography, http://www.physicalgeography.net/fundamentals/9s.html (July 2004).

Shamrma, T., Kiran, V.S., Singh, T.P., Trivedi, A.V., Navalgund, R.R. (2001) Hydrologic response of a watershed to land use changes: a remote sensing and GIS approach, International Journal of Remote Sensing, 22, 11, 2095-2108.

Shekhar, S., Chawla, S. (2003) Spatial databases; a tour. Upper Saddle River, NJ: Pearson Education.

Wollheim, W. M., Pellerin, B.A., Vörösmarty, C.J., Hopkinson, C.S. (2005) N retention in urbanizing headwater catchments in northeastern Massachusetts, Ecosystems, 8:871-884.

Yager, R.R., (1998) Including importances in OWA aggregations using fuzzy systems modeling, IEEE, 6, 2, 286-294.