ESTIMATING NONPOINT SOURCE POLLUTION FOR THE TWIN CITIES METROPOLITAN AREA USING LANDSCAPE VARIABLES

STEVEN M. KLOIBER
Metropolitan Council, 230 East Fifth Street, St. Paul, MN 55101
(e-mail: steve.kloiber@metc.state.mn.us,
Tel.: +(651)-602-1056, Fax: +(651)-602-1130)

(Received 29 August 2005; accepted 3 January 2006)

Abstract. Several stream monitoring programs aim to measure nonpoint source pollutant loading to the major river systems in the Twin Cities Metropolitan Area (TCMA) of Minneapolis and St. Paul, Minnesota. However, due to cost and logistical considerations, only a portion of the total nonpoint source load can be effectively monitored. Regression models were developed relating nonpoint source pollutant yields to landscape characteristics in order to estimate the total nonpoint source contributions of nutrients and suspended sediments to the Mississippi, Minnesota, and St. Croix Rivers from the region. The regression models were generally strong with r-squared values ranging from 0.57 for total suspended solids yield to 0.90 for nitrate yield. The model factors included both land cover variables, such as the percent row crop or the percent urban land as well as soil variables, such as the percent organic matter or the percent clay. These results highlight the importance of considering both land cover and soils when estimating nonpoint source pollutant loads. Using the fitted regression models, the estimated annual average total suspended solids, total phosphorus, nitrate, and total Kjeldahl nitrogen loads for the TCMA are 104,000 metric t/yr, 354 t/yr, 3,580 t/yr, and 1,760 t/yr, respectively.

Keywords: geographic information systems, landscape, modeling, nonpoint, nutrients, sediment, regression

1. Introduction

Like the rest of the United States, the Twin Cities Metropolitan Area (TCMA) of Minneapolis and St. Paul, Minnesota has made major advances in the past 25 years in controlling pollution discharges to rivers and lakes that come from industries and sewage treatment plants (EPA 2000). Unfortunately, efforts to control pollution from diffuse, nonpoint sources (NPS) have not been nearly as effective. The principal pathway of NPS pollution is when rainfall, snowmelt, or irrigation runs over land or through the ground, picks up pollutants, and carries them to rivers and lakes. Taken together, urban and agricultural NPS pollution are the leading sources of water quality impairment of the nation's lakes and rivers (EPA 2002).

To begin to address NPS pollution more effectively, we must first estimate how much NPS pollution reaches a waterbody and where this pollution originates. However, there is considerable difficulty in measuring the spatial distribution because the nature of NPS pollution is that it comes from many diffuse sources that are distributed across the landscape. An effective solution to this problem is to measure

Water, Air, and Soil Pollution (2006) 172: 313–335
DOI: 10.1007/s11270-006-9083-4

pollution at the outlet of a watershed, thus integrating all the pollution sources within a watershed. To this end, the Metropolitan Council established a stream-monitoring program designed to measure pollutant loads coming from watersheds that are tributary to the three major rivers in the TCMA. Many of the tributary streams within the TCMA are monitored through this and other programs; however, cost and logistical issues make it infeasible to measure all NPS contributions to these rivers. The purpose of this paper is to present a summary of annual NPS pollutant loads of nutrients and suspended solids for the monitored watersheds and to develop estimates for the unmonitored watershed.

A variety of methods have been used to estimate nonpoint source loads for unmonitored watersheds. These range from applying simple pollutant loading factors based on monitored watersheds with similar land use (Loehr, 1974) to using complex, semi-mechanistic watershed models such as the Soil and Water Assessment Tool (SWAT) or the Hydrologic Simulation Program FORTRAN (HSPF). The approach used here relies on developing a regression model with which to extrapolate annual pollutant loads for the unmonitored watersheds. This approach is similar to that used by Driver and Tasker (1988) and more recently by Jones *et al.* (2001). The advantage of this approach is that it requires only slightly more effort than the land-use-based loading factor approach and significantly less effort than using a model such as HSPF or SWAT, while still providing reasonably accurate results of annual NPS pollutant load.

1.1. Study area description

The TCMA includes a seven-county region centered on the cities of Minneapolis and St. Paul, MN. The region has a population of 2.64 million people (US Census 2000). Standing out among the region's abundant water resources are the three major rivers that intersect here: the Mississippi, the Minnesota, and St. Croix Rivers. The study area for this analysis includes all land that drains to these three rivers between the three points representing the river inflows and the one point representing the combined river outflow (Figure 1). The inflow points include the Minnesota River at Jordan, MN, the Mississippi River at Anoka, MN, and the St. Croix River at Stillwater, MN. The combined river outflow point occurs at the Mississippi River at Red Wing, MN. The study area covers portions of four major watersheds and encompasses an area of 7,250 square kilometers. The Mississippi – Twin Cities watershed has the most urban development of the four watersheds with 44% of its land between medium-low density urban and high density urban (Table I and Figure 2). The Minnesota – Shakopee watershed is second in terms of urban land at 24% in these same urban land classes, but agricultural land is still predominant with 41% in row crop, small grain, or hay/pasture. The St. Croix – Stillwater watershed and the Mississippi – Rush – Vermillion watershed are mostly agricultural with about 74% and 57% of their land in row crop, small grain, or hay/pasture.

TABLE I
Land cover occurrence by major watershed

Land cover	Minnesota-Shakopee	Mississippi-Rush-Vermillion	Mississippi-Twin Cities	St. Croix-Stillwater
Low density urban	6%	3%	8%	1%
Medium-low density urban	9%	5%	16%	2%
Medium density urban	8%	5%	15%	1%
Medium-high density urban	4%	3%	7%	0%
High density urban	3%	1%	6%	1%
Forest	11%	14%	13%	14%
Turf grass	1%	1%	1%	1%
Water	5%	3%	8%	3%
Wetland (Non-Forested)	7%	2%	8%	1%
Hay/pasture	9%	9%	2%	23%
Row crop	31%	47%	9%	51%
Small grains	1%	1%	0%	0%
Other herbaceous	5%	4%	5%	2%
Total land area (km^2)	1,850	1,120	2,620	1,660

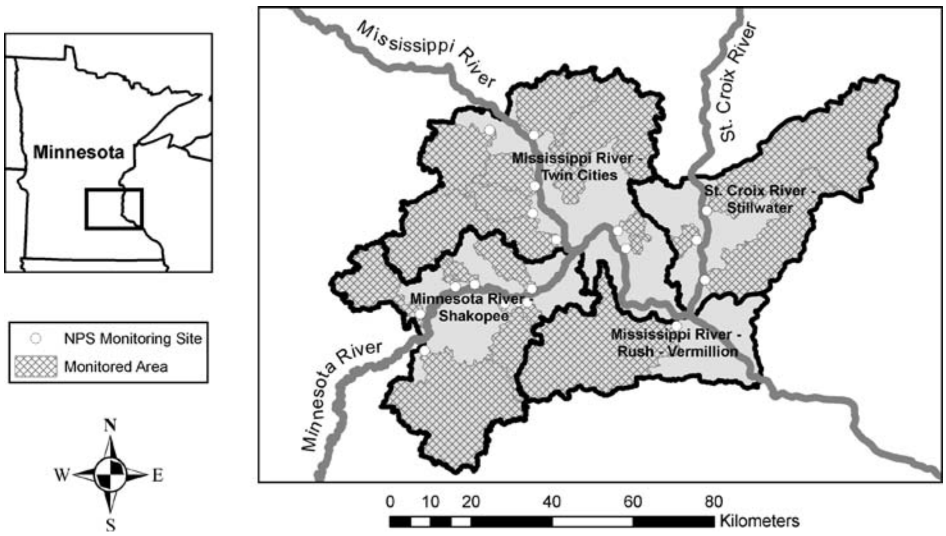

Figure 1. The extent of the study area includes portions of four major watersheds. The portions of the study area that are covered by watershed monitoring are shown with a cross-hatch pattern.

The topography and surficial geology of the TCMA are characterized by the glacial history of the region. A series of glacial episodes has resulted in several major geomorphic associations (Wright, 1972). Each of these associations have distinctive topographic and geological characteristics. For example, in the northern

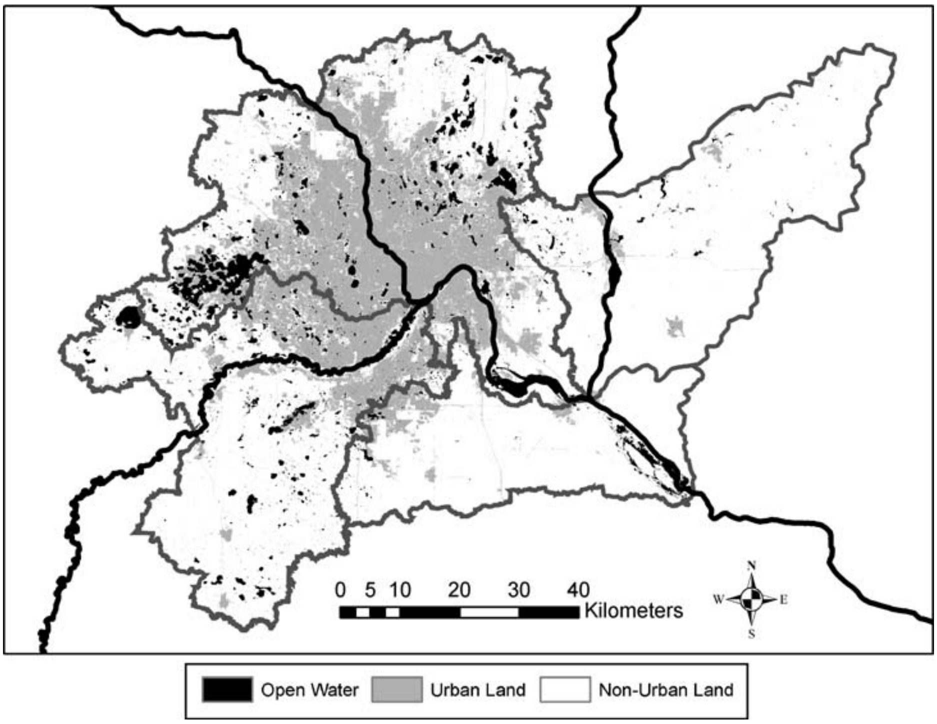

Figure 2. Distribution of urban land cover for the study area per the National Land Cover Dataset. These data were developed and distributed by the U.S. Geological Survey.

part of the study area, the melting waters of the retreating glaciers left flat, sandy outwash plains with numerous shallow wetlands. The central part of the region is characterized by gently rolling hills and many small to medium sized glacial lakes. Sediments range from red sandstone, shale, and agate from the Lake Superior Region to gray and brown clays derived from shales of North Dakota and Canada. The southeastern part of the study area lies beyond the margin of this last glacial period and is comprised of older drift material. The glacial history and subsequent modification of the surficial sediments is reflected in the distribution and character of the soils of the region. As a result, important hydrologic characteristics such as soil permeability and runoff potential have a spatial distribution that reflects this history (Figure 3).

2. Methods

Annual NPS pollution loads to the major rivers from monitored tributaries were calculated from daily average flow and concentration data for water samples collected

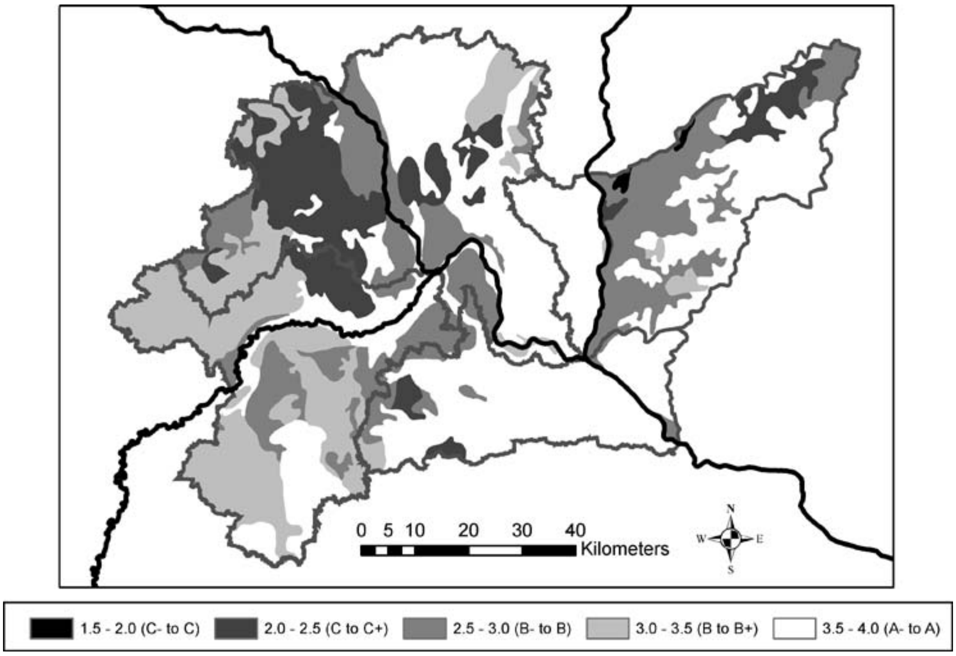

Figure 3. Distribution of soil hydrologic group for the study area per STATSGO Dataset. The data were developed and distributed by the U.S. Natural Resource Conservation Service (NRCS). The NRCS assigns letter codes for soils with differing hydrologic characteristics. These letter codes were converted to numbers for this analysis. Higher numbers indicate more porous soils and potentially less surface runoff.

over a three-year period (2001–2003). The pollutants included in this analysis are total suspended solids (TSS), total phosphorus (TP), total Kjeldahl nitrogen (TKN), and nitrate (NO_3). Forward stepwise, multiple regression was used to estimate NPS pollutant yield for the unmonitored portion of the study area. The data and statistical methods used are briefly summarized here.

2.1. Monitored pollutant data

Data from 24 stream-monitoring stations located in or near the TCMA were used for this analysis (Table II). These sites are generally located near the mouths of streams that are tributary to one of the three major rivers. Twenty of these sites are within the study area (Figure 1). Data from four additional tributaries that are adjacent to the study area were included in this analysis for the purposes of increasing the number of observations used in the regression model, but NPS pollutant loads from these streams are not included in the final NPS load total.

Eighteen monitoring sites are operated under Metropolitan Council monitoring programs. The data from these sites includes continuous 15-minute measurements

TABLE II

Monitored tributary watersheds and pollutant yields

Watershed name	Major watershed	Monitoring program	Area (ha)	Pollutant yield (kg/ha/yr)			
				TSS	TP	NO_3	TKN
Bevens Creek[1]	Minnesota-Shakopee	MCES	33,750	578	1.16	19.78	4.64
Bluff Creek	Minnesota-Shakopee	MCES	1,470	728	0.88	1.68	3.23
Carver Creek	Minnesota-Shakopee	MCES	21,460	454	1.01	5.21	4.39
Credit River	Minnesota-Shakopee	MCES	13,300	159	0.42	2.14	2.00
Eagle Creek	Minnesota-Shakopee	MCES	310	295	1.35	6.19	8.13
Nine Mile Creek	Minnesota-Shakopee	MCES	9,730	210	0.60	1.38	4.30
Riley Creek	Minnesota-Shakopee	MCES	2,730	681	0.71	1.35	3.46
Sand Creek	Minnesota-Shakopee	MCES	60,390	768	1.01	10.29	4.59
Willow Creek	Minnesota-Shakopee	MCES	2,120	122	0.33	1.14	3.36
Bassett Creek	Mississippi-Twin Cities	MCES	10,370	143	0.58	2.04	5.32
Battle Creek	Mississippi-Twin Cities	MCES	2,970	164	0.40	0.97	2.54
Coon Creek	Mississippi-Twin Cities	CCWD	23,430	210	0.57	2.94	
Elm Creek	Mississippi-Twin Cities	USGS	22,140	82	1.04	1.62	3.42
Fish Creek	Mississippi-Twin Cities	MCES	1,320	83	0.53	2.68	2.62
Minnehaha Creek	Mississippi-Twin Cities	MCES	45,730	20	0.13	0.46	1.59
Rice Creek	Mississippi-Twin Cities	RCWD	47,800	88	0.35	0.28	1.63
Shingle Creek	Mississippi-Twin Cities	SCWMC	10,670	94	0.40	1.39	
Browns Creek[1]	St. Croix-Stillwater	MCES	7,410	182	0.36	1.13	2.50
Carnelian-Marine[1]	St. Croix-Stillwater	MCES	7,780	2	0.02	0.10	0.33
Kinnickinnic River	St. Croix-Stillwater	USGS	40,400	32	0.21	10.74	0.77
Silver Creek[1]	St. Croix-Stillwater	MCES	1,930	29	0.07	0.33	0.79
Valley Creek	St. Croix-Stillwater	MCES	3,410	78	0.35	21.37	1.72
Willow River	St. Croix-Stillwater	USGS	71,880	18	0.14	3.91	1.10
Vermillion River	Mississippi-Rush-Vermillion	MCES	69,940	82	0.91	9.01	2.35

[1]These watersheds were adjacent to the NPS study area. They were only included for the purpose of regression modeling.

of flow that are recorded by remote Campbell or ISCO datalogger systems. These data are used to calculate daily average flow for the ice-off season (approximately mid-March through November). Daily flow during the ice-on season is typically estimated from the late fall recession curve, but is generally a very small portion of the annual flow. During storm runoff events, flow-weighted composite samples are collected using automatic samplers for laboratory analysis of a variety of non-point source pollutants including TSS, TP, TKN, and NO_3 (10–15 samples/year). During baseflow conditions, grab samples are obtained for laboratory analysis of water quality variables (12 samples/year). The total number of samples per year is typically between 22 and 27 samples across a range of flow conditions and seasons. Laboratory analysis of water quality samples is performed by the Metropolitan Council's laboratory. Additional details regarding the stream monitoring methods are given elsewhere (Metropolitan Council 2003)

In addition to the eighteen Metropolitan Council monitoring stations used in this analysis, data have been obtained for six other tributaries within the study area by other agencies (Table II). Elm Creek, Willow River, and the Kinnickinnic River were monitored by the USGS. However, the monitoring period for the Willow River and the Kinnickinnic River only covered a one-year period (Lenz *et al.*, 2003). Data for Coon Creek, Rice Creek, and Shingle Creek were all obtained from their respective local watershed management organizations. Shingle Creek and Rice Creek had monitoring data for the selected 2001–2003 study period, but loading estimates for Coon Creek had to rely on a four year period of data from 1995–1998.

Load estimates were obtained from existing sources in most cases. The Metropolitan Council (2004; 2005) previously reported TSS, TP, and NO_3 loads for all streams except for Coon Creek, Shingle Creek, Willow River, and the Kinnickinnic River. All of the loads presented in these reports were calculated using FLUX (Walker, 1996). Loads for Shingle Creek, Willow River, and Kinnickinnic River were obtained from other published reports (Wenck, 2004; Lenz *et al.*, 2003). Pollutant loads for Shingle Creek were calculated using FLUX, while pollutant loads for the Willow and Kinnickinnic were calculated using Estimator (Cohn *et al.*, 1989). Gaps in the pollutant load data were filled in for this study using FLUX. These gaps included TKN loads for the Metropolitan Council monitoring sites and all pollutant loads for Coon Creek. These loads were calculated following the same methodology used previously by the Metropolitan Council and briefly summarized here.

FLUX provides several alternative calculation methods and data stratification schemes. The calculation method used for each pollutant and monitoring station are determined independently from each other, primarily based on three criteria. The difference between the predicted and observed values (i.e. the prediction residuals) should be independent of date, season, and flow to avoid bias in the load estimate. This criterion is evaluated by examining the significance of the slope of a plot of the residuals versus date, month, and flow. The data should be stratified in a manner that ensures that the sampled flows are representative of the total population of the daily flows for each strata. This criterion is evaluated using the Student's T-test to compare the means of the sampled flow and the total flow for each stratum. If these criteria are met, the method providing the lowest coefficient of variation is selected. The pollutant yield was calculated by dividing the mass load per unit time by the watershed area.

An analysis of climate data from 22 stations distributed across the region show that the spatial variability of annual precipitation was small; however, there were inter-annual differences. Annual precipitation values were classified as dry, normal, or wet by comparing the values for each station to the long-term normal precipitation for the region. Observations that fell below the 25th percentile were considered dry, observations falling between the 25th and 75th percentile were considered normal, and observations above the 75th percentile were considered wet. Using this classification system, 2001 was a normal year, 2002 was wet, and 2003 was a dry year. This study did not attempt to address inter-annual differences in nonpoint pollution yield. Using the average pollutant yields from 2001–2003 should minimize the annual differences.

Point source discharges were mapped and evaluated for potentially significant pollutant load contributions. The only major point source contribution to the monitored tributaries was the Empire wastewater treatment plant (WWTP) discharge to the Vermillion River. The contribution of the Empire WWTP was subtracted from the total monitored pollutant load for the Vermillion River prior to calculating the NPS pollutant yield for this watershed. All other major wastewater sources discharge directly to the three major rivers (the Mississippi, the Minnesota, and the St. Croix). A separate accounting of the point source discharges to the major rivers is beyond the scope of the present study.

2.2. Landscape data

The University of Minnesota Remote Sensing and Geospatial Analysis Laboratory (Yuan *et al.*, 2005) developed a 30-meter raster GIS data set of land cover and imperviousness using a multi-temporal classification of Landsat imagery from 2002. Because this data set was limited to just the seven-county area, the National Land Cover Data (NLCD) was used for the outlying portions of the study area. The NLCD data is a 30-meter raster GIS dataset developed by the USGS and the EPA using 1992 Landsat imagery. A few of the NLCD land cover classes were reclassified to be consistent with the TCMA land cover data set. The NLCD was retrieved from the USGS data server (http://seamless.usgs.gov/). The area of each land cover class was tabulated by watershed. Classes that were less than 1% in all watersheds were dropped from the analysis.

Digital soil survey data at the county level (e.g. Soil Survey Geographic Database) are not available for the entire study area; therefore, this analysis relies on State Soil Geographic (STATSGO) database. The STATSGO data for Minnesota and Wisconsin were obtained from the Natural Resources Conservation Service (http://www.ncgc.nrcs.usda.gov). A selection of soil attributes that have the potential to affect water quality and runoff volume were extracted from the STATSGO tables. These attributes include the following variables: hydrologic soil group, organic matter content, clay content, soil erodibility (K factor), soil permeability, calcium carbonate content, cation exchange capacity, and percent of soil passing through a number 40 sieve (<0.5 mm).

All variables are obtained from the soil layer table of STATSGO, with the exception of hydrologic soil group, which comes from the component table. To simplify the analysis, only the upper-most soil layer was considered. These data were joined to their respective soil component tables. Hydrologic soil group is specified as an ordinal alphabetic character in STATSGO, with values ranging from A through D. Soils of hydrologic group A have the highest permeability, while soils of hydrologic group D have the lowest permeability. In order to perform a statistical analysis with this variable, the letter groupings have to be converted to numeric values. This was done by assigning numbers from four for hydrologic soil group "A" to one for hydrologic soil group "D". The hydrologic soil group is sometimes assigned two alternate values (e.g. A/D). These values represent the hydrologic soil group in the drained and undrained condition. These two classes were split with the first value being stored as hydrologic soil group 1 and the second being stored as hydrologic soil group 2. Both classifications were retained for the analysis.

STATSGO contains mapped soil units that actually represent soil associations of several distinct, but related soil types. The relative contribution of each component soil type is specified in the database. To simplify the analysis, the soils data were summarized by map unit using a weighted-average with the component percentage as the weighting factor. The resulting table contains a weighted-average of soil characteristics for each unique map unit. This table was then joined to the spatial data. The soils data were intersected with the watershed boundary data layer, and area-weighted average values were calculated for each watershed unit.

Streams from the National Hydrography Data (NHD), which includes streams that are shown on a 1:24,000 scale map, were used to calculate the percent riparian area. The riparian area was defined as a 100-meter buffer on either side of the stream. Slopes were derived from National Elevation Data (NED) and then classified into three slope ranges: high slope (>5%), medium slope (1–5%), and low slope (<1%). The percent occurrence of slope class was then summarized by watershed.

The geographic coordinate systems of these data sets were reprojected to the Universal Transverse Mercator Zone 15 NAD83 projection using ArcView 9.0 prior to geographic analysis. The data were combined and clipped as necessary to provide complete data coverage for a rectangular region encompassing the study area.

2.3. Statistical analysis methods

Statistical analysis was performed using SPSS (version 10.0) for Windows. Principal component analysis was used prior to developing regression models for pollutant export to explore the structure of the relationships between variables, to characterize the correlation between the independent variables, and to identify a reduced number of landscape factors that may be potentially related to water quality. Factors were extracted using the varimax rotation and only factors with eigenvalues greater than one were retained for further analysis.

Forward, stepwise, multiple regression was used to develop predictive models of pollutant yield for each of the four variables: TSS, TP, TKN, and NO_3 as well as the water yield. The regression equation was limited to the independent variables that were statistically significant ($P<0.05$). The variance inflation factors (VIF) were calculated for each variable to determine if collinearity (correlation between the independent variables) was significant. Any variables with a VIF greater than 3 were excluded. In addition, the Durbin-Watson test was used to detect autocorrelation in the data. The 95% confidence intervals for the predicted values were estimated from the standard error of the Y-estimate.

3. Results

3.1. Monitored NPS pollutant yields and loads

Most (60%) of the monitored TSS pollutant yields were between 59 and 359 kg/ha/yr (Table II). The TSS yield was about 50 times higher than the NO_3 yield, which was most typically between 1.07 and 7.32 kg/ha/yr. NO_3 yields and TKN yields were fairly comparable, with most TKN yields occurring between 1.60 and 4.37 kg/ha/yr. Most TP yields occurred between 0.28 and 0.95 kg/ha/yr, which is about 0.25% to 0.50% of the TSS yield.

The monitored NPS loads of TSS, TP, NO_3, and TKN for the study area are 87,000, 253, 2,390, and 1,040 metric tons per year (t/yr), respectively. The monitored portion of the study area is 462,000 ha and represents 64% of the total area. Simply extrapolating using the average yield for the monitored tributaries results in estimated NPS loads of TSS, TP, NO_3, and TKN of 49,000, 142, 1,340, and 590 t/yr for the unmonitored portion of the study area.

Water yield for the monitored tributaries was most typically (60% of sites) within 16.2 to 27.2 cm/yr. The average annual precipitation at the Minneapolis – St. Paul International Airport for 2001–2003 was 80.0 cm/yr. Eagle Creek had a water yield of 263 cm/yr, which is more than four times the annual precipitation. This is undoubtedly the result of the considerable groundwater discharge that this small stream receives from Boiling Springs, a small pool located near the headwaters of this creek from which groundwater percolates. In fact, the monitored water yield and pollutant loads for Eagle Creek bear little relationship to its watershed characteristics. As such, the yield results for this stream were excluded from the subsequent regression analysis.

The water yield for Valley Creek was 53.5 cm/yr, which is about 67% of the annual precipitation. While not exceeding total precipitation, this water yield is still quite high considering the land cover and soils for this watershed. Like Eagle Creek, Valley Creek receives a significant portion of its water from groundwater. This groundwater inflow allows both streams support cold water fisheries. In the case of Valley Creek, this groundwater input also comes with a high concentration of NO_3 (Zapp and Almendinger, 2001). Due to the significance of the groundwater input, and the uncertainty surrounding how much of the yield could be attributed to watershed processes, Valley Creek was also excluded from subsequent regression analysis.

3.2. Landscape characteristics

One of the important considerations for developing a predictive model is that the sampled population should adequately represent the range for the independent variables for the total population. The measure of the various land cover classes for the monitored watershed ranges 0% to 66% with most land cover classes having a range from 0% to 30% or less (Table III). Some land cover classes, such as turf grass, small grains, and other herbaceous vegetation, have very small ranges. In most cases, the range of the independent variables for the total population is well represented by the sampled population. The exceptions include medium density urban, high density urban, other herbaceous, imperviousness, and calcium carbonate content, which all have maximum values for the population that exceed the maximum values for the sample. In addition, the minimum value of cation exchange capacity for the total population is below the minimum for the sample population. However, calcium carbonate content is the only one of these variables that is used in any of the final regression models and its effect is fairly limited. Therefore, for this analysis the sample range can be considered generally representative of the range for the population.

Having a representative sample population is important for at least two reasons. First, the strength of the correlation depends, in part, on the range of variation for the independent variables. Therefore, having samples that represent the broadest range possible of the independent variables will add to the strength of the model. Second, using a regression model to extrapolate results for watersheds with values for independent variables exceeding the range for the sampled population may be problematic. Doing so requires an assumption that the behavior of the regression model continues beyond the range of independent variables for which it was developed, which may or may not be the case.

3.3. Principal component analysis

The principal component analysis performed on 26 landscape variables produced seven factors with eigenvalues greater than one. A similar analysis on water yield and pollutant yield produced two factors with eigenvalues greater than one. A Pearson correlation matrix between these factors revealed that there were four landscape factors that were significantly correlated to water and pollutant yield (Table IV). The first landscape factor is characterized by high loading on land cover variables with strong positive loading for urban land cover variables and strong negative loading for agricultural variables. The second factor has strong positive loading for soil organic matter content, cation exchange capacity, percent riparian area and low topographic slope, while it has strong negative loading for the soil erodibility K-factor. The third factor appears to represent another set of soil variables with strong positive loading for hydrologic soil group and permeability and strong negative loading for clay and the percent of soil passing through a No.40 sieve. The

TABLE III
Summary of landscape characteristics

Variable	Units	Abbreviation	Min	Max	Mean	CV	Skew
Land Cover							
Low Density Urban	%	LD_URB	0	16	8	0.61	0.14
Medium-Low Density Urban	%	MLD_URB	1	27	13	0.68	0.16
Medium Density Urban	%	MD_URB	0	27	11	0.85	0.46
Medium-High Density Urban	%	MHD_URB	0	16	5	0.98	0.58
High Density Urban	%	HD_URB	0	11	4	0.89	0.81
Forest	%	FOREST	3	41	15	0.60	1.26
Turf Grass	%	TURF	0	4	1	0.94	**1.52**
Water	%	WATER	0	18	5	0.99	**1.62**
Wetlands (Non-Forested)	%	WETLAND	0	16	6	0.68	0.65
Hay/Pasture	%	HAY_PAST	0	29	7	1.09	**1.88**
Row Crop	%	ROW	0	66	21	1.05	0.84
Small Grains	%	GRAIN	0	2	1	0.68	1.01
Other Herbaceous	%	HERB	0	9	5	0.62	0.13
Imperviousness	%	IMPRV	0	43	17	0.84	0.41
Drainage and Topography							
Riparian Area	%	RIPARIAN	4	24	14	0.37	0.04
Low Slope Area	%	LOW_SLP	14	77	35	0.47	0.94
High Slope Area	%	HIGH_SLP	0	29	13	0.64	0.56
Soils							
Calcium Carbonate Content	%	$CACO_3$	0.00	1.00	0.19	**1.40**	**1.94**
Cation Exchange Capacity	meq/100 g	CEC	7.32	52.3	20.7	0.61	0.94
Clay Content	%	CLAY	3.74	23.5	13.8	0.36	0.22
Hydrologic Soil Group 1	Unitless	HYDGRP1	2.70	3.89	2.99	0.09	**1.77**
Hydrologic Soil Group 2	Unitless	HYDGRP2	2.07	3.32	2.55	0.13	0.32
Erodibility Factor	tons/ac	KFFACT	0.12	0.31	0.24	0.17	0.73
Sieve No.40	%	NO40	61.9	83.3	72.4	0.09	0.12
Organic Matter	%	OM	2.63	25.08	9.52	0.63	0.85
Permeability	in/hr	PERM	1.22	9.98	3.16	0.63	**2.06**

Bolded entries indicate where the spread is large (Coefficient of Variation >1.2) or where the distribution is significantly skewed (Skewness Coefficient >1.5). A highly skewed distribution will have most of its values clustered at one end with few values at one extreme.

final landscape factor that was significantly correlated to water and pollutant yield was actually the seventh factor extracted, meaning that, among retained factors, it explains the least amount of variability in the original set of landscape data. Interpretation of this factor is somewhat unclear. It has strong positive loading for the percent of land cover in grain crops, the soil calcium carbonate content, the

TABLE IV
Correlation results among landscape factors and water quality and quantity factors

	Water quality/Quantity factor 1		Water quality/Quantity factor 2	
Landscape factor	Correlation coefficient	p-value	Correlation coefficient	*p*-value
1	−0.280	0.232	**0.677**	**0.001**
2	**0.482**	**0.031**	0.075	0.753
3	−0.468	0.037	**−0.513**	**0.021**
4	0.011	0.964	0.009	0.969
5	0.255	0.278	0.426	0.061
6	−0.235	0.318	−0.074	0.755
7	**0.465**	**0.039**	−0.102	0.669

Bolded entries indicate significant correlations.

soil clay content, and the percent of land cover in sand and gravel extraction. From these results, it can be expected that the regression models for water and pollutant export may include one land use variable and possibly two soil variables.

A K-means cluster analysis was performed on these four landscape factors with six clusters specified a priori. This procedure minimizes the variability within each cluster and maximizes the variability between clusters. The clusters suggest a spatial pattern with tributaries located in the St. Croix-Stillwater watershed clustering together and the western tributaries of the Minnesota-Shakopee watershed in another cluster (Figure 4). Tributaries in the Mississippi-Twin Cities watershed and two others from the eastern portion of the Minnesota-Shakopee watershed formed two clusters. The Vermillion River stood alone in another cluster. This geographic distribution roughly follows the pattern of geomorphic associations and suggests that there are spatial patterns of pollutant export across the region.

3.4. Regression Model Results

Statistically significant regression models were developed for all four pollutant variables and water yield (Table V). The value of r^2 ranges from 0.57 for TSS yield to 0.90 for NO_3 yield. Each equation includes two or three independent variables and each equation also includes a combination of land cover and soil variables. Soil variables accounted for 57% to 87% of the r-squared value for all the independent variables except for NO_3 yield, for which 90% of the r-squared values was attributable to the percentage of row crop land cover in the watershed. This fact suggests that soil characteristics can be quite important in explaining the variability of nonpoint source yields for at least some pollutants..

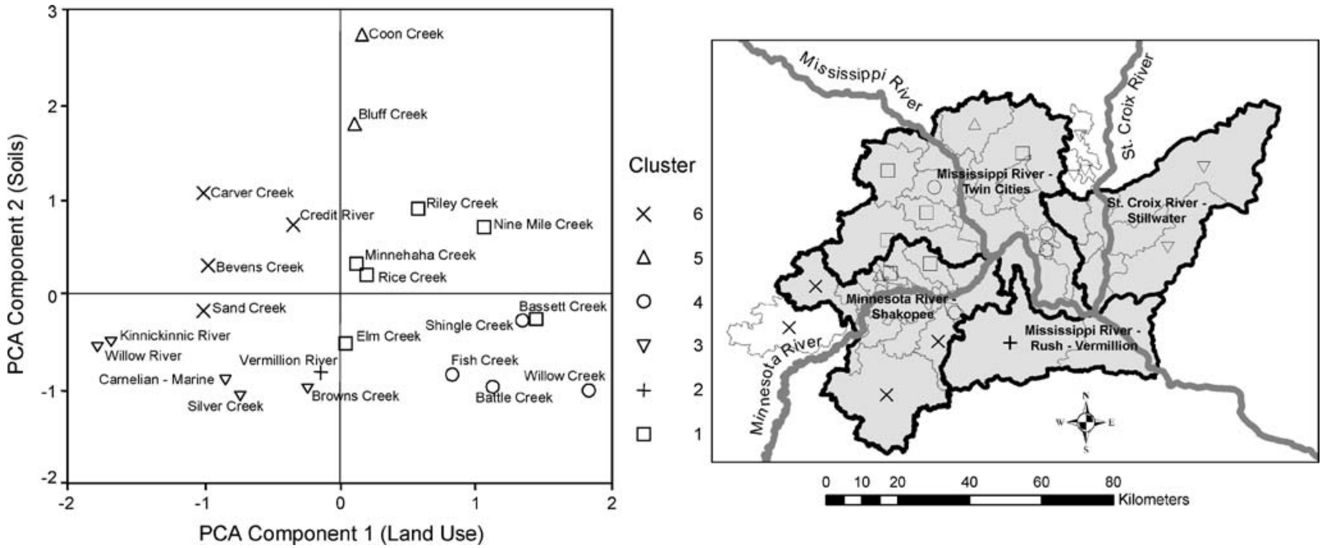

Figure 4. Scatterplot of the first two principal components for the landscape characteristic data. The *x*-axis is related to land use while the y-axis is related to soil characteristics. The symbols represent clusters of watersheds with similar characteristics. The geographic distribution of the cluster is depicted on the map.

TABLE V
Regression modeling results for yield

Regression equation	Adjusted r^2	Standard error of estimate
Water = 0.776(MLD_URB) + 0.703(NO40) − 12.1(HYDGRP1) − 3.54	0.73	4.000
ln(TSS) = 0.156(OM) − 0.135(WATER) − 0.246(PERM) + 4.545	0.57	0.944
TP = 0.0415(CLAY) + 0.0217(OM) − 0.0209(WATER) − 0.158	0.69	0.189
TKN = 0.227(CLAY) + 0.132(MHD_URB) − 1.22	0.65	0.859
$NO_3 = 3.88\times10^{-5}(ROW)^3 + 6.39(CACO_3) + 0.555$	0.90	1.533

Dependent variable units are kg/ha/yr for all variables except water, which has units of cm/yr.

The variance inflation factors are all below a value of 2, indicating that collinearity is not an issue. All independent variables shown in these equations were significant at a level of $P < 0.05$. A graphical review of the dependent variable plotted versus the adjusted predicted value shows that the residuals are well behaved (Figure 5). However, it is apparent that the observed distribution for NO_3 yield is rather skewed, which may be somewhat problematic because the upper end of the observed range of NO_3 yield is only represented by a few data points. The leverage that these few points exert leads to additional uncertainty regarding the regression model. Nevertheless, the regression models appear to be reasonable.

The regression models were then used to estimate the pollutant yields for the unmonitored watersheds. In the case of TSS yield, a simple bias-correction factor was applied to account for the reverse transformation bias (Gilbert 1987). Pollutant load estimates for the unmonitored portion of the study area are 17,000 t/yr, 101 t/yr, 1190 t/yr, and 720 t/yr for TSS, TP, NO_3, and TKN. The prediction residuals were used to calculate relative confidence intervals ($p < 0.05$). Based on this analysis, the estimated uncertainties for the unmonitored NPS pollutant load for TSS, TP, NO_3, and TKN are ±40%, ±14%, ±16%, and ±12%, respectively. The total NPS load was determined by adding these estimated pollutant loads for the unmonitored area to the monitored loads. The total annual average TSS, TP, NO_3, and TKN loads for the TCMA are 104,000 metric t/yr, 354 t/yr, 3,580 t/yr, and 1,760 t/yr, respectively (Table VI).

Looking at the spatial distribution of pollutant yield it is clear to see that higher TSS yields occur in the Minnesota-Shakopee watershed located in the southwest TCMA (Figure 6). The five highest monitored TSS yields occur in Sand Creek, Bluff Creek, Riley Creek, Bevens Creek, and Carver Creek. All of these watersheds are in the Minnesota-Shakopee watershed. The yields for these watersheds range from 454 to 768 kg/ha/yr. The TSS yields for all the other monitored tributary watersheds range from 2 to 295 kg/ha/yr. The lowest TSS yields tend to occur in the St. Croix-Stillwater watershed. With the exception of Brown's Creek, the

TABLE VI
Summary of NPS pollutant loads by major watershed

	TSS (t/yr)	TP (t/yr)	NO_3 (t/yr)	TKN (t/yr)
Minnesota-Shakopee	74,300	142	1,100	712
Mississippi-Twin Cities	18,100	100	367	586
St. Croix-Stillwater	4,500	30	969	185
Mississippi-Rush-Vermillion	7,020	82	1,140	274
Total NPS Load	104,000	354	3,580	1,760

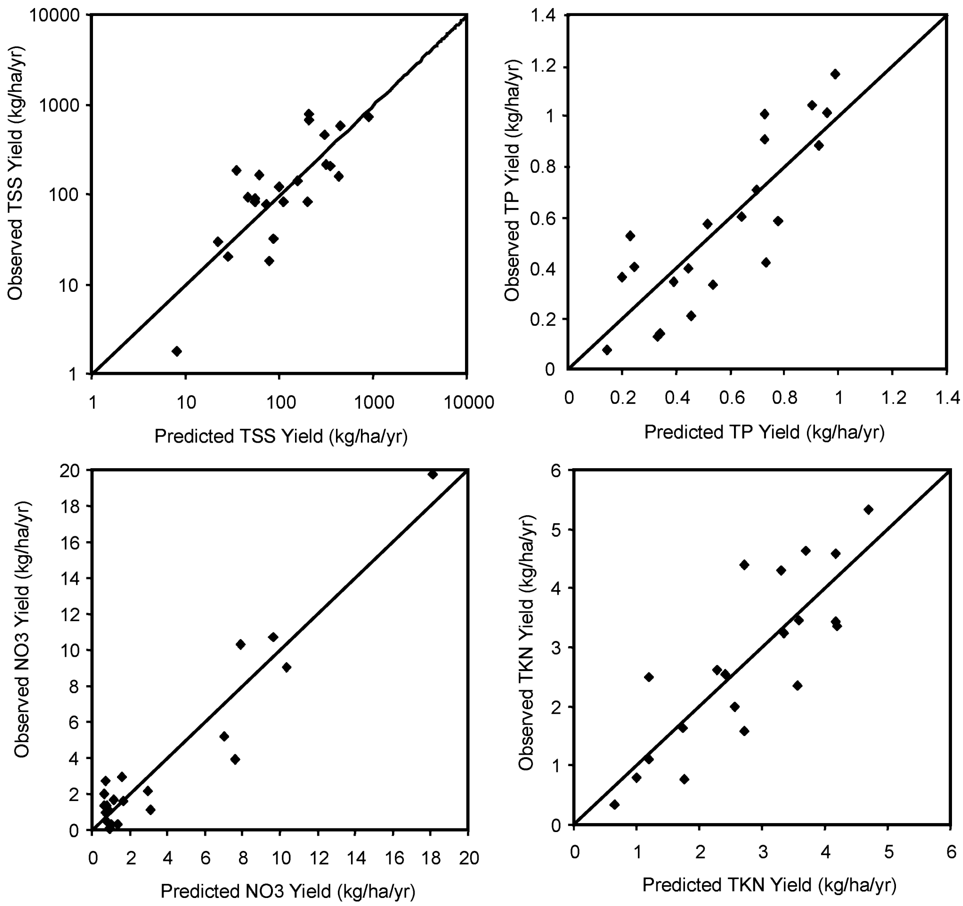

Figure 5. Scatterplots of TSS, TP, NO_3, and TKN show good agreement between the predicted and observed values.

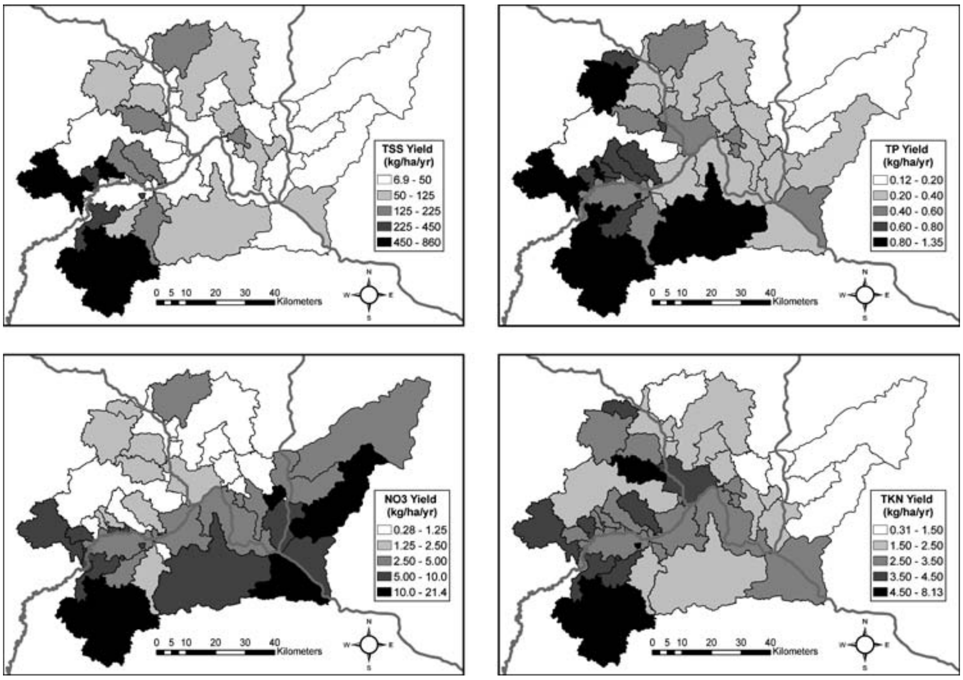

Figure 6. Average annual yield for (a) TSS, (b) TP, (c) NO_3, and (d) TKN in units of (kg/ha/yr). Monitored yields were used where available. Yields estimated using from the regression models were used for unmonitored watersheds.

tributary watersheds in the St. Croix-Stillwater watershed have TSS yields that range from 2 to 78 kg/ha/yr. In terms of loading, the Minnesota-Shakopee watershed contributes 72% of the total estimated TSS load from NPS in the study area. The TSS load from the Mississippi-Twin Cities watershed constitutes 17% of the total.

Four of the five lowest monitored TP yields were also found for tributaries in the St. Croix-Stillwater watershed. About 40% of the total NPS TP load comes from the Minnesota-Shakopee watershed and another 28% comes from the Mississippi-Twin Cities watershed. Similarly, six of the seven highest monitored TKN yields were observed in tributaries of the Minnesota-Shakopee watershed. The proportions of the NPS TKN load that are contributed by the Minnesota-Shakopee and Mississippi-Twin Cities watersheds are 40% and 33%, respectively. Thus, it appears that there is a general gradient of pollutant yield for TSS, TP, and TKN across the study area, with lower yields in the northeast and higher yields in the southwest (Figures 6a, 6b, and 6c).

The spatial pattern for NO_3 yield is significantly different from the pattern for the other three pollutants (Figure 6d). Three out of the highest seven monitored NO_3 yields were observed in tributaries of the St. Croix-Stillwater watershed. Another three out of the highest seven NO_3 yield observations were found in the Minnesota-Shakopee watershed and one came from the Mississippi-Rush-Vermillion watershed. With the exception of Eagle Creek, the monitored watersheds with highest NO_3 yield also had the most agricultural land cover. The NPS load contributions of NO_3 for the Mississippi-Rush-Vermillion, the Minnesota-Shakopee, and the St. Croix-Stillwater watershed were 32%, 31%, and 27%, respectively. The most urbanized major watershed, the Mississippi-Twin Cities, only contributed about 10% of the estimated total NO_3 load from NPS.

4. Discusssion

4.1. Yields

The monitored ranges of watershed yields are wide. The ranges for TKN and TP yields span more than an order of magnitude, while the ranges for TSS and NO_3 span more than two orders of magnitude. This characteristic is reflected in the range of yield values reported by Loehr (1974) and Langland *et al.* (1998) as well as the range of concentration values reported by the Nationwide Urban Runoff Program (EPA 1983). The range of pollutant yields found in this study were similar to what has been reported in other studies of urban and agricultural watersheds, such as Langland *et al.* (1988), Loehr (1974), and Mulcahy (1990).

The range of TSS yields for the monitored watersheds in this study was 2 kg/ha/yr to 768 kg/ha/yr. This appears to be substantially lower than the range reported for Chesapeake Bay tributaries by Langland *et al.* (1998) of 32.5 kg/ha/yr to 1,570 kg/ha/yr. However, these ranges are affected by a few extreme values. If we compare the 20th and 80th percentiles, we find that the yields for the TCMA were between 59 and 359 kg/ha/yr and those for the Chesapeake Bay were between 78 and 466 kg/ha/yr, which are comparable. Comparing the 20th and 80th percentiles for TP yield, we find that the TCMA yields typically range between 0.28 and 0.95 kg/ha/yr, while yields for the Chesapeake Bay tributaries range between 0.39 to 0.90 kg/ha/yr. These ranges are quite comparable to each other. Also, comparing the 20th and 80th percentiles for NO_3 yield, the typical yields for the TCMA and Chesapeake Bay area are comparable with the typical range for the TCMA between 1.07 and 7.32 and the typical range for the Chesapeake Bay between 1.28 and 6.30 kg/ha/yr.

4.2. Loads

The regression model estimates of pollutant load for the unmonitored portion of the study area are 17,000 t/yr, 101 t/yr, 1,190 t/yr, and 720 t/yr for TSS, TP, NO_3, and TKN. These estimates can be compared to the loads estimated from simply taking the average monitored yields and extrapolating to the unmonitored area. The model estimated load of TSS for the unmonitored area is 65% lower than the estimate obtained by simple extrapolation (49,000 t/yr). The model estimate of TP load is 29% lower than the simple extrapolation estimate (142 t/yr) and the model estimate of TKN load is 22% higher than the extrapolated estimate (590 t/yr). The model estimate for NO_3 load is 11% lower than the extrapolated estimate (1,340 t/yr). Considering the confidence intervals about the model estimates that were presented earlier, the modeling results are significantly different from a simple extrapolation for TSS, TP, and TKN. Therefore, the modeling effort appears to be warranted for these parameters. The simple extrapolation estimate for NO_3 load was within the uncertainty estimate of the model, suggesting that in this case the simple estimate may suffice.

4.3. Model strength

Jones *et al.* (2001) reported r-square values for similar regression models of total ammonia yield and total NO_3 yield of 0.65 to 0.86, respectively, which are comparable to the r-square for TKN yield for this study of 0.65 and the r-square for NO_3 yield of 0.90. In the case of NO_3 yield, agricultural land cover is the dominant independent variable for both the studies. Jones *et al.* (2001) also reported an r-square value for suspended sediment of 0.79. For this study, the r-square value for the TSS yield regression model was weaker at 0.57. However, this study found an r-square value for TP yield regression model of 0.69, which was stronger than the value reported by Jones *et al.* (2001) of 0.45. Driver and Tasker (1988) reported a range of r-square values of 0.20 to 0.65 for estimating mean seasonal or annual loads. Thus, the regression models are generally strong as measured by the adjusted r-square values and compare favorably to other such published values.

4.4. Model factors

The strongest factors that are included in the models (p-values < 0.001) can be broadly classified into three categories; land cover factors, soil and geology factors, and transport and retention factors. Land cover factors include medium-low density urban land, medium-high density urban land, and row cropped land. The soil and geology factors include the percent of soil finer than 0.5 mm, percent clay, percent organic matter, and the calcium carbonate content. The pollutant transportation and retention factor included in two of the models is the percent of open water in the watershed.

Human landscape alteration of agricultural and urban land use has long been cited as a cause of increased non-point source pollutant yield (Loehr 1974; EPA 1983). Results from this study suggest that agricultural land use is a dominant factor affecting nitrate yields. Similarly, a study in eastern Wisconsin conducted by Robertson and Saad (1996) found concentrations of dissolved nitrate were highest in agricultural areas, moderate in urban areas, and lowest in forested areas.

Lenz *et al.* (2003) found that there was no relationship between the percentage of agricultural area and the annual nutrient and sediment yields for St. Croix River tributaries. However, the authors noted that they only had data from one year and that there was considerable spatial variation in the precipitation patterns that may have obscured any such relationship. They also noted that when comparing samples collected at baseflow, agricultural tributaries did have higher concentrations of nitrogen compounds.

Land use was also found to be significant for two other predictive models: water yield and TKN. Urban land use was significantly correlated to water yield. In this case, the percentage of land in medium-low density urban development had the strongest correlation to water yield; however, many of the other urban land use variables, such as the total imperviousness, were nearly as strongly related. The relationship between imperviousness and runoff volume is well known (Schueler, 1994) and is one of the most important calibration variables for urban runoff models such as the Stormwater Management Model (Huber and Dickinson 1988; Warwick and Tadepalli, 1991). The other relationship that included an urban land use variable was the model for TKN yield. The percentage of medium-high density urban land (60–80% imperviousness) was positively correlated with TKN yield. However, the p-value for this variable was greater than that for the percent clay content of the soil, suggesting that land use has less effect on TKN yield than the soil character.

Although urban and agricultural land use clearly have a relationship to pollutant yield, it is interesting that they don't appear in the predictive relationships for TP and TSS. One might reasonably expect that water yield and pollutant yield are correlated and that because urban land use is a significant factor in predicting water yield that this factor should also occur in the regression equations for the other pollutant yield variables. However, TKN yield is the only pollutant yield that is significantly correlated with water yield. One possible explanation for this behavior is that the concentrations of the other pollutants decrease as urbanization and water yield increase, thus obscuring this relationship. Looking at the monitored concentrations across an agricultural-urban gradient it is clear that this is the case for TSS and NO_3 and possibly for TP as well although the decrease in concentration is not quite as dramatic.

Looking at the remaining explanatory variables, it is clear that the role of soils and geology in understanding the nonpoint pollutant yield should not be overlooked. These results show that more than half of the explanatory power of the regression models was attributable to soil variables, except for the case of NO_3 yield. Acknowledgement of this role is reflected in the ecoregion concept presented by Omernik (1987) or the agro-ecoregions of Birr and Mulla (2002). This approach seeks to group land areas into classes according to the pattern of associated landscape characteristics including bedrock geology, soils, vegetation, and land surface form as well as land use. These regions are then used as a means of grouping streams and lakes for the purpose of better describing water quality. In this approach, the within class water quality variability is reduced, while the between class variability is increased. Although the categorical approach used for these studies is different than the continuous regression modeling method used in this study, both approaches support the notion that soils and geology are significantly correlated to surface water quality and nonpoint source pollutant yield.

Two soil variables are included in the regression model for water yield. Water yield increases with increases in the percent of soils passing through a No. 40 sieve ($<$0.5mm) and it decreases with increases in the weighted soil hydrologic group. This is not surprising given the fact that soil variables such as these have long been included in runoff models such as the Green-Ampt model and the Soil Conservation Service model (Chow *et al.*, 1988).

Clay content shows up as a significant variable in the regression models for TP and TKN, while organic matter content shows up as a variable in the TSS and TP models. Whether these variables have a causal relationship with pollutant yield is unclear. Furthermore, Robertson and Saad (1996), reported TP concentrations in the surface water of eastern Wisconsin were significantly higher in areas with sandy soils as opposed to clayey soils, which appears to contradict the findings presented in this paper. They also reported that there were no significant differences in surface water concentrations of nitrate plus nitrite (NO_x) or TKN between areas of different surficial deposit texture or bedrock type. However, there are some key methodological differences between these studies. Robertson and Saad (1996) were evaluating pollutant concentrations and not pollutant yields and they also used a categorical approach with a higher level of generalization for soils and geology. Regardless of the potential discrepancies, both studies show that surface water quality and nonpoint source pollution are related to soil and geology as well as with land use.

In addition, several variables included in the analysis have potential relationships to watershed retention of pollutants. These include the percentage of open water, the percentage of wetland area, the percentage of low slope land, and the percentage of riparian area. One of these variables, the percentage of open water, a significant variable for the TSS and TP regression models. The regression model shows that pollutant yields of TSS and TP are inversely related to the percentage of open water. Put another way, the yield of TSS and TP decreases as the percentage of open water area increases, which suggests a possible mechanistic relationship based on sedimentation and retention of particulate pollutants in the lakes and ponds of these watersheds. Driscoll (1986) described a similar relationship of pollutant retention in a dimensionless analysis of the effect of detention ponds on urban runoff quality. This study showed that as pond area increased, expressed as a percent of the total watershed area, that TSS removal increased.

The regression modeling approach presented here was found to be an effective means of extrapolating results from monitored watersheds to unmonitored watersheds. Other approaches may also be useful, but these should account for the importance of soils, geology and water impoundment as well as land use in assessing nonpoint source pollution. Rather than simply defining pollutant yields by land use category, as is frequently done, a better approach would be to define pollutant yields for ecoregions or agro-ecoregions, which incorporate these other factors.

Acknowledgements

The stream monitoring data were collected by the Metropolitan Council and several local cooperating agencies. The pollutant load estimates for the monitored streams were calculated by Karen Jensen, Cassie Champion, and Leigh Harrod as well as by the author. I would also like to thank Karen Jensen and Marcel Jouseau of the Metropolitan Council as well as Larry Baker of the University of Minnesota Water Resource Center for providing review comments.

References

Birr, A. S. and Mulla, D. J.: 2002, 'Relationship between lake and groundwater quality patterns and Minnesota agroecoregions', *Hydrological Sci. Tech.* **18**(1–4), 31–41.

Chow, V. T., Maidment, D. R. and Mays, L. W.: 1988, *Applied Hydrology*. McGraw-Hill Book Company. New York, NY. pp. 572.

Cohn, T. A., Delong, L. L., Gilroy, E. J., Hirsch, R. M. and Wells, D. K.: 1989, 'Estimating constituent loads'. *Water Res. Research*. **25**(5): 937–942.

Driscoll, E. D.: 1986, 'Detention and retention controls for urban runoff', in B. Urbonas and L. A. Roesner (eds), *Urban Runoff Quality – Impact and Quality Enhancement*, ASCE. pp. 381–393.

Driver, N. E. and Tasker, G. D.: 1988, 'Techniques for Estimation of Storm-Runoff Loads, Volumes, and Selected Constituent Concentrations in Urban Watersheds in the United States. U.S. Geological Survey', *Water Supply Paper* 2363. pp. 51.

EPA: 1983, *Results of the Nationwide Urban Runoff Program: Volume 1 – Final Report*. United States Environmental Protection Agency, Office of Water. Washington, DC. (EPA-832-R-83-112). pp. 200.

EPA: 2000, *Progress in Water Quality, 'An Evaluation of the National Investment in Municipal Wastewater Treatment': Chapter 12 – Upper Mississippi River Case Study*. United States Environmental Protection Agency, Office of Water. Washington, DC. (EPA-832-R-00-008). pp. 34.

EPA: 2002, *National Water Quality Inventory: 2000 Report*. United States Environmental Protection Agency, Office of Water. Washington, DC. (EPA-841-R-02-001). pp. 207.

Huber, W. C. and Dickinson, R. E.: 1988, *Storm Water Management Model Version 4 User's Manual*. United States Environmental Protection Agency, Athens, GA. (EPA/600/3-88/001a). pp. 595.

Jones, K. B., Neale, A. C., Nash, M. S., Van Remortel, R. D., Wickham, J. D., Riitters, R. H. and O'Neill, R. V.: 2001, 'Predicting nutrient and sediment loadings to streams from landscape metrics': A multiple watershed study from the United States Mid-Atlantic Region', *Landscape Ecology*, **16**(4): 301–312.

Langland, M. J., Edwards, R. E. and Darrell, L. C.: 1998, *Status Yields and Trends of Nutrient and Sediment Data and Methods of Analysis for the Non-Tidal Data Collection Programs. Chesapeake Bay Drainage Basin: 1985–1996*. U.S. Geological Survey, Baltimore, MD. Open File Report 98–17. pp. 67.

Lenz, B. N., Robertson, D. M., Fallon, J. D. and Ferrin, R.: 2003, *Concentrations and Loads, and Benthic Invertebrate Data for Tributaries to the St. Croix River, Wisconsin and Minnesota, 1997–99*. U.S. Geological Survey. Water Resources Investigation Report 01-4162. pp. 70.

Loehr, R. C.: 1974, 'Characteristics and comparative magnitude of non-point sources', *J Water Poll. Control Fed.* **46**(8). 1849–1872.

Metropolitan Council: 2003, *Quality Assurance Program Plan: Stream Monitoring*. Metropolitan Council, Environmental Services Division. St. Paul, MN. pp. 28.

Metropolitan Council: 2004, *2002 Stream Monitoring and Assessment for 11 Metropolitan Area Streams*. Metropolitan Council, Environmental Services Division. St. Paul, MN. pp. 171.

Metropolitan Council: 2005, *2003 Stream Monitoring Report*. Metropolitan Council, Environmental Services Division. St. Paul, MN. pp. 288.

Mulcahy, J. P.: 1990, *Phosphorus Export in the Twin Cities Metropolitan Area*. Metropolitan Council,. St. Paul, MN. pp. 26.

Omernik, J. M.: 1987, 'Ecoregions of the conterminous United States', *Ann. Assoc. Am. Geog*. **77**(1), 118–125.

Robertson, D. M. and Saad, D. A.: 1996, *Water Quality Assessment of the Western Lake Michigan Drainages–Analysis of Available Information on Nutrients and Suspended Sediment, Water Years 1971–1990*. U.S. Geological Survey Water-Resources Investigations Report 96–4012. pp. 167.

Walker, W. W.: 1996, *Simplified Procedures for Eutrophication Assessment and Prediction: User Manual*. U.S. Army Corp of Engineers, Waterways Experiment Station, Vicksburg, MS. Instruction Report W-96-2. pp. 238.

Warwick, J. J. and Tadepalli, P.: 1991, 'Efficacy of SWMM Application', *J. Water Resour. Plng. and Mgmt*. **117**(3), 352–366.

Wenck Associates, Inc: 2004, 'Shingle Creek and West Mississippi Watershed Management Commissions': Annual Water Quality Report 2003. Maple Plain, MN. pp. 36.

Wright, H. E., Jr.: 1972, 'Physiography of Minnesota, in Sims, P. K., and Morey, G. B., (eds), *Geology of Minnesota': A Centennial Volume*. Minnesota Geological Survey. pp. 561–578.

Yuan, F., Sawaya, K. E., Loeffelholz, B. C. and Bauer, M. E.: 2005, 'Landcover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing', *Remote Sens. Environ.* **98**(2–3), 317–328.

Zapp, M. L. and Almendinger, J. E.: 2001, Nutrient Dynamics and Water Quality of Valley Creek, a High-Quality Trout Stream in Southeastern Washington County, Final Project Report to the Valley Branch Watershed District. St. Croix Watershed Research Station, Marine on St. Croix, MN. pp. 31.