Hierarchical Approaches to the Study of Water Quality in Rivers

Spatial scale and terrestrial processes are important in developing models to translate research results to management practices

Carolyn T. Hunsaker and Daniel A. Levine

Land-use change may be the single greatest factor affecting ecological resources. Allan and Flecker (1993), who have identified six major factors threatening the destruction of river ecosystems, state that various transformations of the landscape—hydrologic changes to streams and rivers resulting from changes in land use, habitat alteration, and nonpoint source pollution—are probably the most widespread and potent threats to the well-being of lotic ecosystems.

Measures of landscape structure are necessary to monitor change and assess the risks it poses to ecological resources (Graham et al. 1991, Hunsaker et al. 1990). Landscape ecologists seek to better understand the relationships between landscape structure and ecosystem processes at various spatial scales (Forman and Godron 1986, Risser 1987, Turner 1987, 1989). We use the word *structure* to refer to the spatial relationships of ecosystem characteristics such as vegetation, animal distributions, and soil types. *Processes* or *function* refers to the interactions—that is, the flow of energy, materials, and organisms—between the spatial elements. Because landscapes are spatially heterogeneous areas, or environmental mosaics, the structure and function of landscapes are themselves scale-dependent.

Carolyn T. Hunsaker is an environmental scientist and research staff member in the Environmental Sciences Division of Oak Ridge National Laboratory, Oak Ridge, TN 37831-6038. Daniel A. Levine is an environmental scientist at the Automated Sciences Group, Inc., Oak Ridge, TN 37830.

The location of various types of land use in a watershed is critical to modeling

Scientists often organize spatial scale in a hierarchical manner such as nested watersheds and ecoregions (Crowley 1967, O'Neill et al. 1989). For example, the US Geological Survey (USGS) has defined the Hydrologic Unit Codes as a four-level hierarchy or logical arrangement of river basins where each larger unit is an aggregate of smaller units (Seaber et al. 1984). We are developing methods to characterize landscape attributes that influence water quality at various spatial scales. Understanding how scale, both data resolution and geographic extent, influences landscape characterization and how terrestrial processes affect water quality are critically important for model development and translation of research results from experimental watersheds to management of large drainage basins.

Landscape characterization and water quality

Streams and rivers serve as integrators of terrestrial landscape characteristics and as recipients of pollutants from both the atmosphere and the landscape; thus, large rivers are especially good indicators of cumulative impacts. Many studies have shown that the proportion of different land uses within a watershed can account for some of the variability in river water quality (DelRegno and Atkinson 1988, Omernik 1977, Reckhow et al. 1980, Sivertun et al. 1988). For example, the water quality in a watershed with 50% agricultural land use and an intact forest

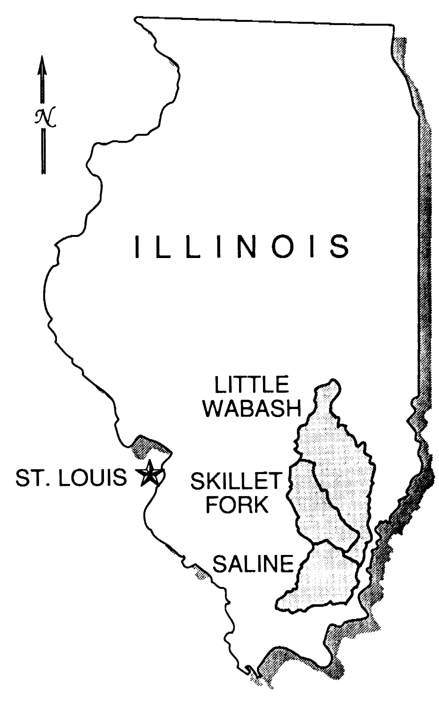

Figure 1. The Wabash River study site in Illinois contains three river basins—the Little Wabash, Skillet Fork, and Saline—comprising 1.35 million ha.

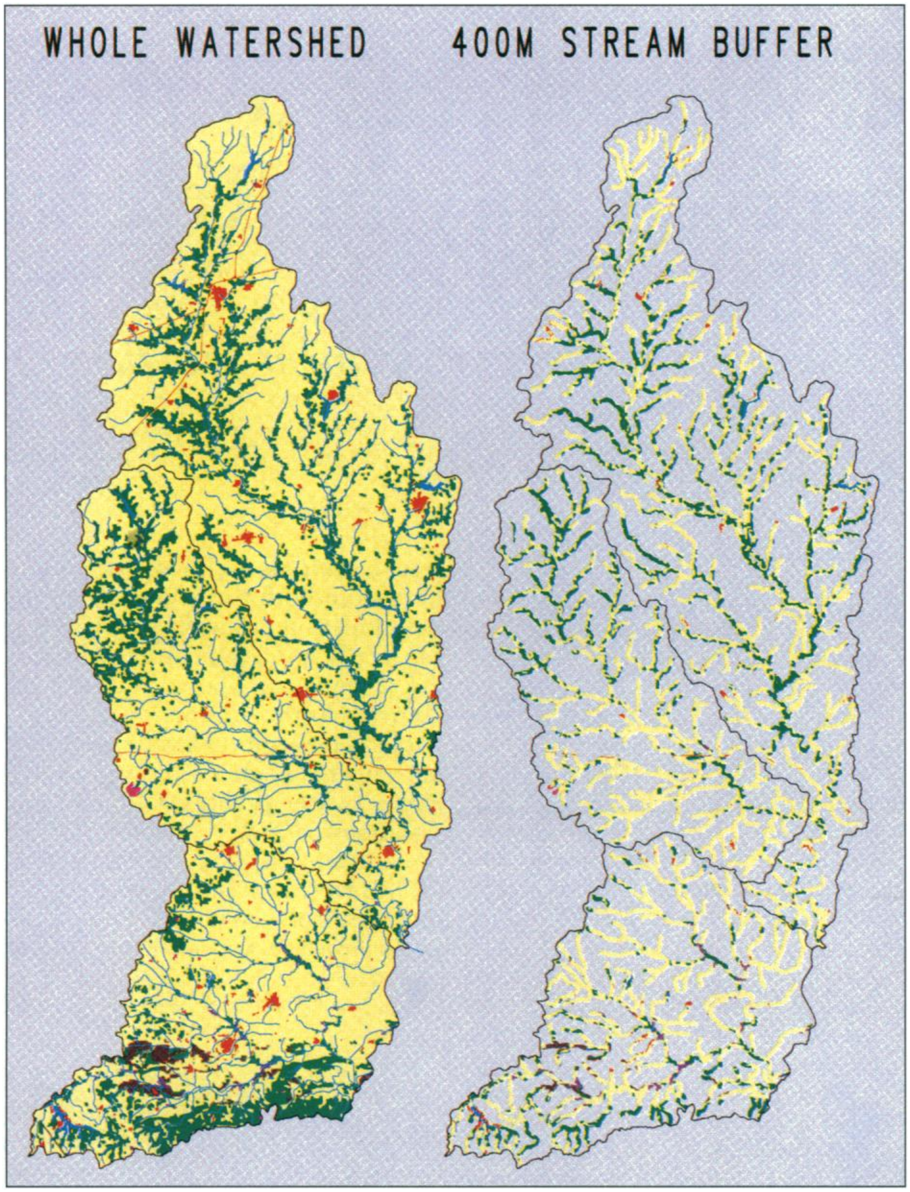

Figure 2. Land-use coverage for the 400-meter equidistant corridor around streams (right) and for the entire Illinois site (left). Key to colors: yellow = agriculture, green = forest, red = urban, blue = water, purple = wetlands, orange = rangeland, brown = barren land.

riparian zone may be expected to be better (e.g., lower turbidity and nutrients) than that in a similar watershed without any riparian zone. In addition, several researchers have addressed the issue of whether land use close to streams is a better predictor of water quality than land use over the entire watershed (Omernik 1981, Osborne and Wiley 1988, Wilkin and Jackson 1983). Research on nutrient and sediment movement within small watersheds with forest or grass buffer areas between streams and disturbed uplands generally supports such statements (Cooper et al. 1987, Lowrance et al. 1984, Peterjohn and Correll 1984, Schlosser and Karl 1981). However, conclusions by Omernik et al. (1981) for larger watersheds from a wide variety of hydrologic settings suggest that upland land uses are as important as near-stream land uses.

As the issues being addressed by ecological research and resource management have become more complex and integrative, an interesting question has surfaced with regard to the spatial construct we use to characterize regions. Geographers have struggled for a long time with spatial interrelationships and questions of geographic characterization. A regional characterization scheme is especially difficult for aquatic ecosystems.

Two accepted approaches exist, watersheds (Likens et al. 1977, Lotspeich 1980) and ecoregions (Omernik 1987), and both can be hierarchically constructed. A watershed is an area of land draining to a specific point on a stream or to a lake or wetland; watersheds are based on topography and the observation that water flows downslope because of gravity. An ecoregion is an area of relative homogeneity based on one or more attributes such as climate, bedrock geology, or soil properties.

Using nested watersheds as the hierarchical regional characterization scheme, this article addresses three questions relevant to characterizing landscape attributes important to water quality:

- Are both the proportions of land uses and the spatial pattern of land uses important for characterizing and modeling river water quality in watersheds of different areas?
- Can land use near the stream better account for the variability in water quality than land use for the entire watershed?
- Does the size of the watershed influence statistical relationships between landscape characteristics and water quality or model performance?

The results of this work are likely to aid the understanding and management of nonpoint-source pollution for large geographic areas.

We performed spatial analyses on raster, or cell, data using several different geographic information systems (which are often referred to by the initials GIS). A geographic information system is a computerized mapping system for capture, storage, management, analysis, and display of spatial and descriptive data. In a raster-based system, numeric values for map data are represented in a grid containing rows and columns of cells of a prescribed size. Each cell corresponds to a fixed area on the earth.

To demonstrate hierarchical approaches to the study of large rivers, we used two case studies: the upper Little Wabash River in Illinois and the Lake Ray Roberts drainage of the Trinity River in Texas. All available monitoring and research data were used. Because we relied on data collected by others and used various sampling designs, we were constrained in the type and intensity of statistical analyses that could be applied. (This problem, which is standard in regional analyses of water quality, grows larger with spatial scale.)

We sought to work in two very different landscapes. Although much water-quality monitoring is done in the United States, few large regions have consistent and long-term monitoring networks that can support a hierarchical analysis. We selected the Illinois and Texas study areas because the water-quality monitoring was adequate for our purposes. In addition, they had land-cover and soils data available in a digital format.

The studies were conducted concurrently; preliminary findings were shared between the projects and helped shape the research. Although the studies undertook two different modeling approaches, lumped for Illinois and spatially explicit for Texas, the use of a hydrologically active area (Novotnoy and Chesters 1981) defined by the stream network and/or topography was incorporated into both. Within a watershed the areas that produce surface runoff are called hydrologically active; the rest of the watershed contributes only to interflow and base flow. The hydrologic activity of an area is a stochastic phenomenon depending on the magnitude and intensity of the storm, soil conditions, and surface characteristics of the area.

The Wabash River study site in southeastern Illinois

The Wabash River study site contains three river basins—Little Wabash, Skillet Fork, and Saline—comprising 1.35 million ha (Figure 1). Land-use/land-cover data came from the USGS (Anderson et al. 1976) based on aerial photographs taken from 1974 to 1976. For land cover (Figure 2), we used a cell size of 200 m on a side (an area of 4 ha). Seven land-use classes occur on the site: agriculture, forest, rangeland, barren, wetland, urban, and water. The stream network and digital elevation models were based on topographic maps with a scale of 1:100,000. Each original cell in the digital elevation model represented 90 m x 70 m; these data were resampled to match the land-cover resolution. One cell in the resampled digital elevation models represented approximately four original cells.

Water chemistry data (as concentrations) were retrieved from the US Environmental Protection Agency's STORET (storage and retrieval) database. We identified 47 water-quality monitoring stations with adequate data in the study area (Figure 3). Stations with fewer than three samples in a year or with variances equal to zero were not used. For each station, data were averaged over the period 1974 through 1977. The mean values were weighted by using the number of samples that went into each mean divided by the variance around the mean. In effect, means with larger sample sizes and lower variances had higher weights than means with smaller sample sizes and larger variances. Only data for total nitrogen, total phosphorus, and conductivity were considered extensive enough for analysis (i.e., enough stations and enough samples per station per year). Not all chemistry parameters were monitored at all stations: total nitrogen was measured at 36 stations, total phosphorus at 33, and conductivity at 36. The watersheds of each station included in our study were delineated on USGS 1:24,000-scale topographic maps, and their boundaries were digitized. The hierarchical structure of the watersheds is illustrated in the network diagram in Figure 3.

Landscape characterization. Several metrics have been proposed to quantify landscape pattern (Baker and Cai 1992, O'Neill et al. 1988). In our study, we included the proportion of the seven land-use types and the amount of edge between different land uses—forest and agriculture, forest and barren, wetland and agriculture, and wetland and barren. We also used the dominance metric, which measures the extent to which one or a few land uses dominate the landscape. Another measure employed was a three-cell contagion to measure the extent to which the landscape is fragmented (Hunsaker et al. 1994). The landscape metrics were calculated using a custom program (Timmins and Hunsaker in press).

Because each of the 13 landscape metrics was needed for each of 47 watersheds and for the corridors around streams, a geographic information system was employed. Individual watershed boundaries, representing the areas draining to each sampling station, were used as an overlay to quantify the total area of each land-cover type within each watershed. Two methods were used to define hydrologically active areas (Figure 4). Equidistance corridors or buffers around the stream network were generated using the ARC/INFO™ geographic information system (ESRI 1987) in widths of 200 m and 400 m (one and two cells) on each side of the stream (Figure 2). These were the smallest areas possible, given the data resolution. A model of the hydrologically active area was also calculated from a digital elevation model using the COUNT program developed by Jensen and Dominque (1988). This program determines the upslope area that drains into each cell. The value in each cell of the resulting file is an actual count of the number of cells that are topographically upslope and therefore contribute hydrologically to that cell. We used an arbitrarily selected COUNT threshold of 35 cells (140 ha) to define the hydrologically active area. Every cell with a value of 35 or greater was assumed to be contributing flow and therefore influencing water quality (i.e., a cell had to have 35 or more cells flowing into it to be considered part of the hydrologically active area). Landscape metrics were calculated for the entire watershed and both of the areas estimated to be hydrologically active using different techniques.

Modeling approach. The spatial analysis capabilities of a geographic information system and multivariate statistics were used to develop empirical/statistical models to address our first two questions. Total nitrogen, total phosphorus, and conductivity were treated as the dependent variables, and the 13 landscape metrics were used as explanatory variables. For the Illinois site, the best reduced model (seven or eight variables) was determined using the *R*-square procedure (SAS Institute 1985). Separate models were developed for the full watersheds, the hydrologically active area based on topography within watersheds (defined by the COUNT program), and the equidistance areas around the stream network within watersheds.

Three generally recommended criteria (Draper and Smith 1981) were used in selecting the best partial model. First, the sum of squared error for the partial model could not differ significantly from the sum of squared error for the full model (13 variables). Second, the amount of variability (R^2) explained by the partial model had to be at least 95% of that explained by the full model. Third, it was preferred that the Cp statistic be positive and reasonably close to the degrees of freedom plus one (Draper and Smith 1981). However, the latter condition was often hard to meet and was relaxed in some cases. Finally, the independent variables in the model had to be scientifically sensible. For example, nutrient levels would logically increase with increased agriculture in a watershed and decrease as the amount of undisturbed forest increases. The effect of collinearity was evaluated using the maximum condition index (Belsley et al. 1980). For all cases, the maximum condition index was less than 13, which indicates only moderate collinearity. Collinearity is a measure of similarity in the variability explained by two parameters and is a measure of independence.

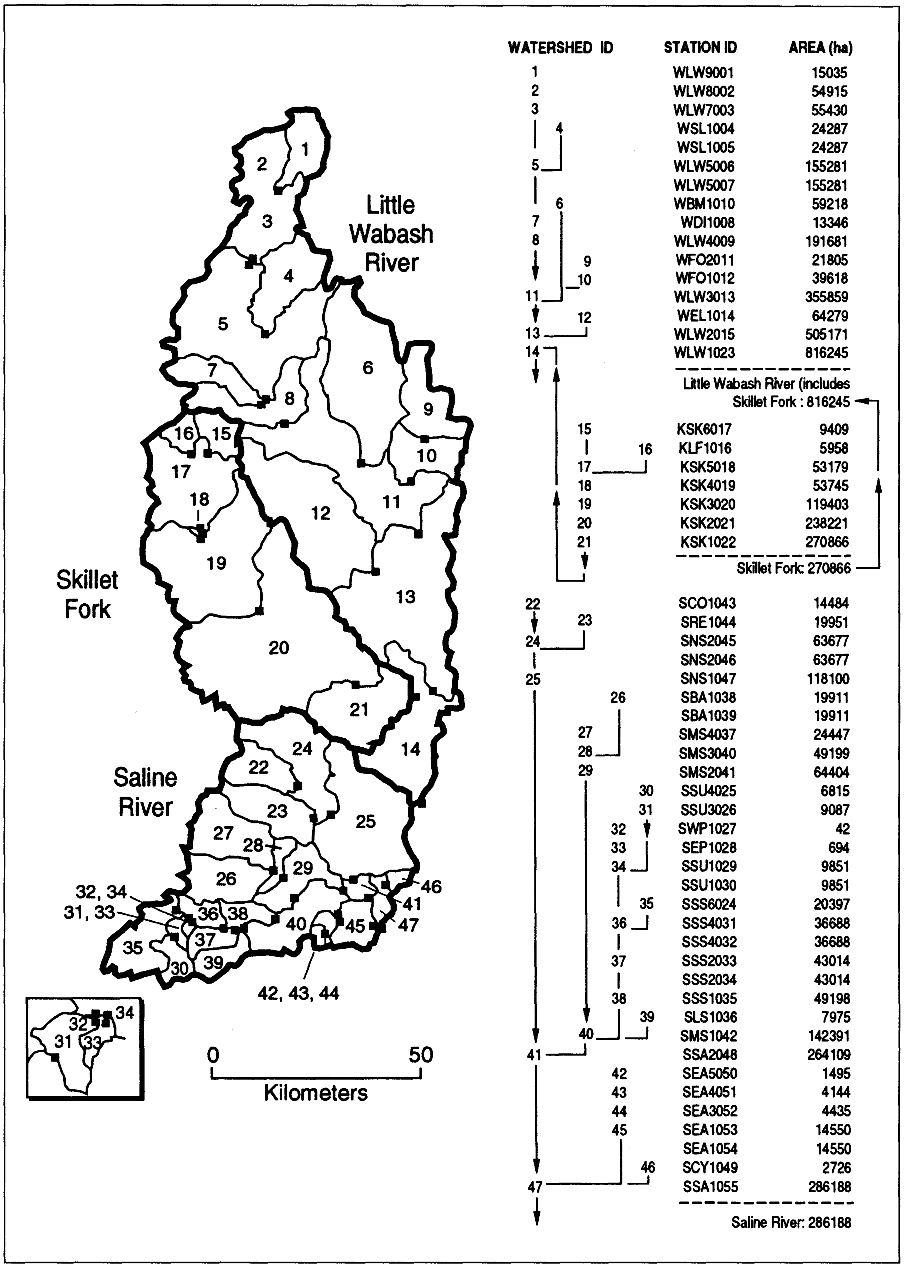

Figure 3. Location of monitoring stations used in the Wabash River study. The chart shows the hydrologic flow between watersheds and the area of each watershed. If two different agencies monitor the same site, only one set of data was used in model development.

Landscape pattern and its relationship to water quality. We evaluated three chemistry parameters that have extensive data collection—total nitrogen, total phosphorus, and conductivity—and that are representative of general water quality. Nitrogen and phosphorus are nutrients that in excess can cause eutrophication in the form of algal blooms and oxygen depletion. Conductivity provides a rapid assessment of the dissolved solids content (ion concentration) of water and thus can serve as a general indicator of water purity.

Relationships between fish and phytoplankton productivity and total dissolved solids concentrations have been developed using the morphoedaphic index (Adams et al. 1983). The regression models for total phosphorus, total nitrogen, and conductivity are presented in Table 1. The R^2 values are given for both the reduced models (seven or eight variables) and for the individual contribution to the variance by the most influential variables. In other words, the R^2 values under the "single variable model" headings are for each explanatory variable regressed by itself against the water-quality parameter.

Both the proportions of land-cover types and their spatial pattern (i.e., amount of edge, dominance, and contagion) are useful in characterizing water quality; however, proportions of cover types consistently account for the most variance (i.e., large R^2). Such landscape metrics can account for 40% to 86% of the variance in stream quality across a range of watershed sizes (1000 to 1.35 million ha). Landscape metrics were least effective for total nitrogen, often accounting for only half of the variance in this parameter. No strong pattern emerged where a few of the same landscape metrics consistently accounted for a significant amount of variance, either within the models for an individual water-quality parameter or within the type of watershed model (entire watershed or hydrologically active area).

Using landscape data for the entire watershed consistently explains the most variance in water quality. Results for equidistance corridors, both 200 m and 400 m, and hydrologically active area based on topography are similar; R^2 values were usually ten units less for those areas than those for the entire watershed. In general, hydrologically active areas around streams contain a significant proportion of the forest remaining in the watersheds. This case is especially true for the Little Wabash basin (Figure 2). The Skillet Fork basin has a predominance of agriculture, even within the stream corridors, while the Saline basin contains a significant amount of barren land (from mining operations) and wetlands near streams.

In general, the direction of correlations was logical and consistent between the models used. For example, disturbed land covers like agriculture, barren, and rangeland have positive associations with water-quality parameters; that is, as the proportion of agriculture increases, so does the amount of nitrogen, phosphorus, and conductivity. Contagion and proportion of forest are negatively correlated with water-quality parameters (Hunsaker et al. 1992). Thus, an area that has contiguous land covers (is not fragmented) or that is dominated by forests tends to have better water quality.

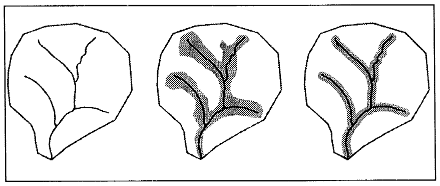

Figure 4. Representation of hydrologically active areas. (left) A stream network within a small watershed. (center) The shaded area represents an irregularly shaped, hydrologically active area surrounding the stream network, which might be produced by an analysis based on topography, soils, and land cover. (right) Shaded area represents a regularly shaped hydrologically active area, which would be produced by an equidistance buffer around the stream network.

The Lake Ray Roberts study site in northern Texas

The Lake Ray Roberts study site comprises 179,821 ha and contains two river systems—the Elm Fork of the Trinity River and the Isle du Bois Creek—which form the headwaters of the Trinity River (Figure 5). These two basins were subdivided into 12 watersheds. The drainage basins of each river system are different in morphometry, land use, and soil type. These differences are delineated by three physiographic regions. The Grand Prairie region, which is drained by the Elm Fork of the Trinity River, is defined by gentle topography, mostly clay loam soils, and cropland and rangeland as the dominant land cover (Figure 6). The Cross Timbers region, which is drained by the western arm of the Isle du Bois River system, is characterized by slightly more severe topography than the other regions, well-drained sandy soils, and forest and pasture as the dominant land-cover types. The Blackland Prairie region, drained by the eastern arm of the Isle du Bois River system, is relatively flat, dominated by clay soils, and has a mixture of cropland and pasture land-cover types.

Water-quality data for this study were obtained from the University of North Texas (Pillard 1988) and were collected biweekly from May 1985 through December 1986. The water-quality parameters used were total phosphorus, total nitrogen, total suspended solids, and instantaneous flow velocity. Total mass loading for a sampling period for each pollutant was determined by multiplying the measured pollutant concentration by the flow for the same date and then multiplying that number by the number of days for the period around each sampling date. Period total loads were then summed to get an annual mass load. Flow and chemistry data were collected during several storm events, and these data dominate the calculations of annual loads.

A cell-based GIS dataset was compiled from various sources. Watershed boundaries, streams, and elevation contours were digitized from 1:24,000 USGS quadrangle maps. Soils data were obtained from the Soil Conservation Service's Map Image Analysis and Display database with a cell resolution of 250 m on a side. The study area contains 105 soil types, and soil attributes were taken from county soil surveys (including mean particle diameter, permeability, and the potential soil erodibility k-factor).

Land-use/land-cover data were obtained from the 1986 Landsat multispectral sensor coverage, classified at 80-meter resolution by the University of North Texas. The land-cover classes used for our modeling were: commercial/residential, industrial/transportation (characterized as in Figure 6c), water, cropland, maintained pasture, natural rangeland, forest, and barren land. A digital elevation model, 20-meter resolution, was developed from the digitized contours and used to calculate slope angle, direction of flow,

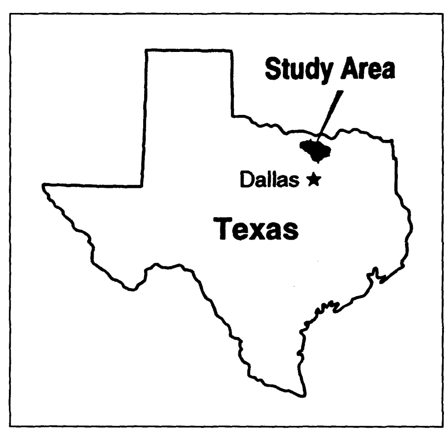

Figure 5. Location of the Lake Ray Roberts watershed (arrowhead).

Table 1. Linear regression models for the Illinois site for entire watershed, contributing area, and 200-meter and 400-meter equidistance corridors. R^2 values are given for both the best seven or eight variable models and for individual independent variables. Variables are a pattern metric, the proportions of watershed in a listed land use or the amounts of edge between two listed land uses.

	Total phosphorus		Total nitrogen		Conductivity	
	Land use	R^2	Land use	R^2	Land use	R^2
Entire watershed						
Single variable model	Urban	25	Urban	17	Barren	55
	Barren	14	Wetland/agriculture	16	Rangeland	37
	Forest/barren*	12	Wetland	11	Forest/barren	24
	Wetland/barren	12	Forest/agriculture	7	Wetland/agriculture	18
	Rangeland	11	Forest	3	Wetland	14
Best complex model	(8 variables)	86	(8 variables)	53	(8 variables)	84
Hydrologically active area						
Single variable model	Contagion†	19	Urban	8	Rangeland	48
	Barren	11	Forest	6	Contagion	11
	Rangeland	11	Barren	5	Barren	7
	Wetland	5	Rangeland	5	Forest	7
	Wetland/barren	4	Wetland/agriculture	4	Forest/barren	5
Best complex model	(8 variables)	76	(8 variables)	41	(8 variables)	76
200-meter and 400-meter corridors‡						
Single variable model	Barren		Contagion		Barren	
	Wetland/agriculture		Dominance†		Forest/barren	
	Forest/barren				Rangeland	
	Contagion				Forest	
Best complex model	(8 variables)	71	(8 variables)	42	(7 variables)	75

*Amount of edge between forest and barren land covers. All variables with a "/" are amount of edge.
†Contagion and dominance have no units.
‡200-meter and 400-meter corridors were analyzed separately. The variables listed under each water-quality parameter were important in both corridor widths. The order of importance changed slightly and the R^2 changed.

and a COUNT file. Soils and land-use layers were resampled to yield a 20-meter cell size to match the digital elevation model resolution and fit the model requirements.

Modeling approach. The modeling approach for the Ray Roberts study linked statistical modeling of nutrient and sediment delivery from each cell with the hydrologic flowpath for each cell to determine the total nutrient and sediment load that reached each watershed outlet (Figure 7). Levine et al. (1993) reviewed the literature on vegetated filter-strip research and, using consistent criteria, they selected 13 studies for development of statistical models for delivery ratios. A delivery ratio is determined by dividing the amount of sediment or nutrient that flows out of a plot of land by the amount that flows into it. The physical and chemical parameters within the plot of land, among other parameters, are related to the magnitude of the delivery ratio. Both linear (forward variable selection procedure, R-square) and nonlinear (NLIN, Marquardt method) regression models were developed using data from the 13 studies to determine these statistical relationships (SAS Institute 1985). The models predict the amount (in mass or concentration) of total phosphorus, total nitrogen, and total suspended solids delivered from one side of a plot of land to the other. The amount delivered generally depended on soil type, slope, and vegetative cover.

Although the linear models describe the data used to develop them (R^2 values of 0.78 to 90), when they were applied in the overall modeling approach they did not perform well (Levine et al. 1993). Therefore, nonlinear models were developed and incorporated into a geographic information system (IDRISI™), and delivery ratios for each pollutant and for each cell were calculated. The nonlinear equations developed from the vegetated filter strip data used three or four of the following watershed characteristics: distance of flow, soil permeability, Manning's roughness coefficient (represents land cover), soil mean particle diameter, and slope angle. Manning's roughness coefficient was used in the total phosphorus and total nitrogen models but not in the model for total suspended solids.

Delivery ratios were calculated for all cells in each watershed (Figure 6). Delivery ratios were then accumulated along the hydrologic flow path for each cell, based on the flow direction at each cell, to calculate total flow path delivery ratios.

Table 2. Total annual loads for each Texas watershed calculated for the entire watershed and for delivery model results.

Location		Total annual loads calculated from entire watersheds: Total phosphorus (kg/yr)	Total nitrogen (kg/yr)	Total suspended solids (kg/yr)	Total annual loads calculated from delivery models: Total phosphorus (kg/yr)	Total nitrogen (kg/yr)	Total suspended solids (kg/yr)
Cross Timbers watersheds							
Timber Creek	Observed	2.78×10^3	1.67×10^4	9.37×10^5	2.78×10^3	1.67×10^4	9.37×10^5
	Estimated	1.29×10^4	5.94×10^4	6.23×10^6	2.75×10^3	1.55×10^4	1.31×10^6
	% Difference	364	256	564	-1	-7	39
Indian Creek	Observed	2.68×10^3	1.68×10^4	1.37×10^6	2.68×10^3	1.68×10^4	1.37×10^6
	Estimated	1.13×10^4	5.78×10^4	6.51×10^6	2.64×10^3	1.97×10^4	1.86×10^6
	% Difference	321	244	375	-1	17	36
Wolf Creek	Observed	1.27×10^2	2.69×10^3	7.80×10^4	1.27×10^2	2.69×10^3	7.80×10^4
	Estimated	4.93×10^3	2.23×10^4	2.53×10^6	1.05×10^3	6.73×10^3	6.07×10^5
	% Difference	3782	729	3243	726	150	701
IDB1	Observed	4.83×10^4	3.04×10^5	2.56×10^7	4.83×10^4	3.04×10^5	2.56×10^7
	Estimated	1.08×10^5	4.76×10^5	7.34×10^7	3.82×10^4	1.88×10^5	2.62×10^7
	% Difference	123	57	187	-21	-38	2
Grand Prairie watersheds							
Spring Creek	Observed	8.86×10^3	5.13×10^4	2.49×10^6	8.86×10^3	5.13×10^4	2.49×10^6
	Estimated	2.54×10^4	1.06×10^5	1.88×10^7	8.86×10^3	3.64×10^4	6.42×10^6
	% Difference	187	107	655	0	-29	158
TR4	Observed	2.03×10^4	9.76×10^4	1.73×10^6	2.03×10^4	9.76×10^4	1.73×10^6
	Estimated	6.98×10^4	2.93×10^5	1.07×10^8	2.24×10^4	9.60×10^4	4.62×10^7
	% Difference	244	200	6085	10	-2	167
TR3	Observed	2.72×10^4	1.52×10^5	4.33×10^6	2.72×10^4	1.52×10^5	4.33×10^6
	Estimated	9.93×10^4	4.20×10^5	1.74×10^8	2.67×10^4	1.15×10^5	4.98×10^7
	% Difference	265	176	3918	-2	-24	1050
TR2	Observed	3.36×10^4	2.24×10^5	1.13×10^7	3.36×10^4	2.24×10^5	1.13×10^7
	Estimated	1.08×10^5	4.57×10^5	1.80×10^8	2.95×10^4	1.29×10^5	5.19×10^7
	% Difference	221	104	1493	-12	-43	359
TR1	Observed	3.78×10^4	2.78×10^5	1.30×10^7	3.78×10^4	2.78×10^5	1.30×10^7
	Estimated	1.43×10^5	6.03×10^5	2.07×10^8	4.15×10^4	1.79×10^5	5.83×10^7
	% Difference	278	117	1492	21	-35	348
Blackland Prairie watersheds							
Buck Creek	Observed	8.55×10^3	3.89×10^4	3.64×10^6	8.55×10^3	3.89×10^4	3.64×10^6
	Estimated	1.57×10^4	6.49×10^4	1.16×10^7	8.45×10^3	3.79×10^4	6.07×10^7
	% Difference	84	67	219	-1	-3	1567
IDB3	Observed	2.32×10^4	1.62×10^5	1.44×10^7	2.32×10^4	1.62×10^5	1.44×10^7
	Estimated	5.65×10^4	2.47×10^5	3.92×10^7	2.55×10^4	1.21×10^5	1.72×10^7
	% Difference	144	52	172	9	-25	19
IDB2	Observed	3.33×10^4	2.12×10^5	1.85×10^7	3.33×10^4	2.12×10^5	1.85×10^7
	Estimated	7.35×10^4	3.18×10^5	5.15×10^7	3.41×10^4	1.59×10^5	2.34×10^7
	% Difference	121	50	178	2	-25	26

The delivery ratio is assumed to be 1.0 (100% delivery) in stream cells.

The models were calibrated by increasing the density of the stream network using the COUNT file. The COUNT threshold was reduced to allow a smaller area to generate stream flow or 100% delivery. Because the soils, topography, and land-cover types in each physiographic region were different, we expected watersheds to behave differently in the hydrologic sense. Therefore, one watershed in each region was used to calibrate the models, resulting in three different calibration COUNT thresholds for each physiographic region. For each pollutant, the data layer for total flow path delivery ratio was multiplied by the data layer for potential nutrient and sediment loading, resulting in a layer representing total pollutant from each cell that reached

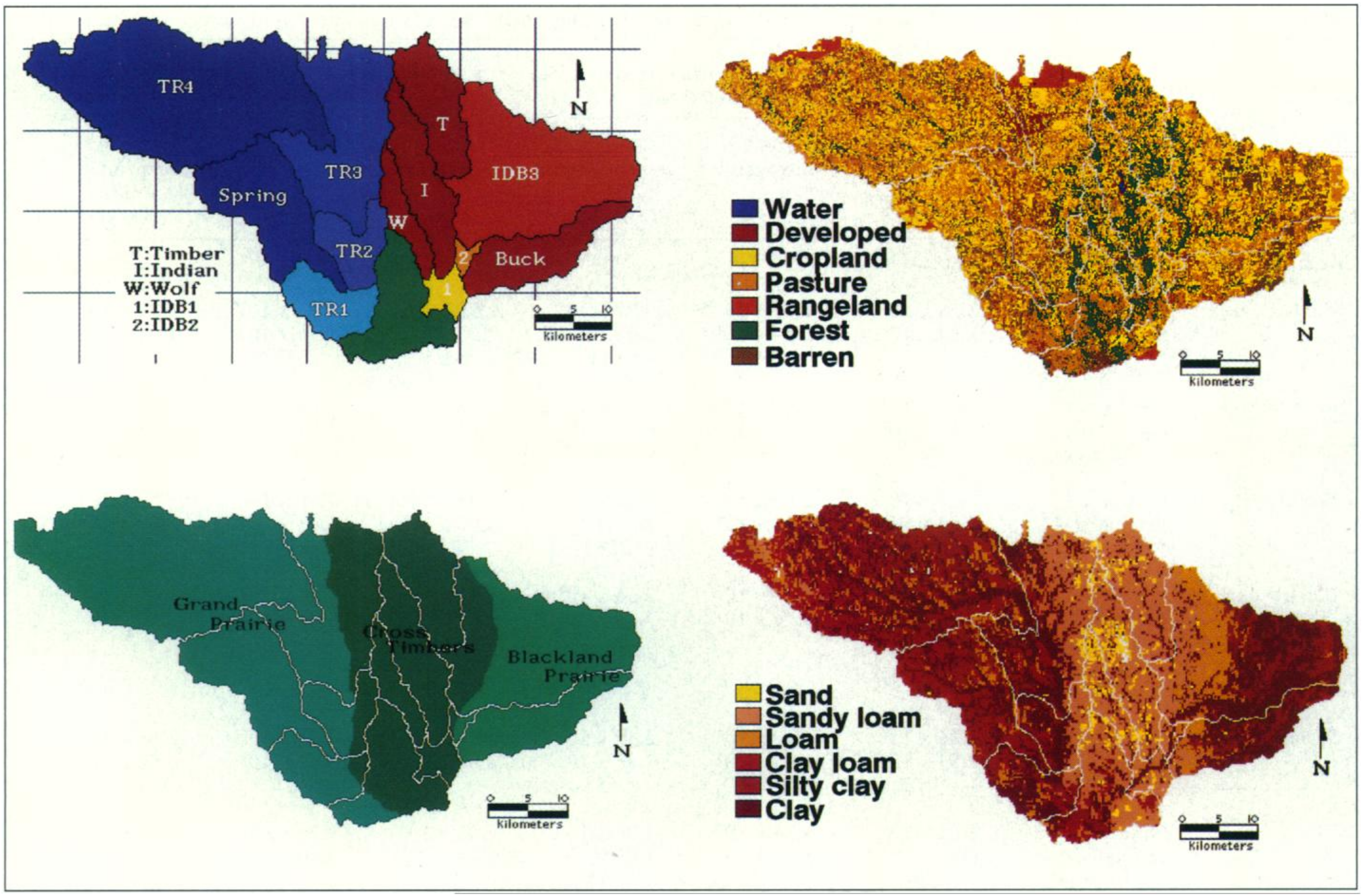

Figure 6. Lake Ray Roberts Basin characteristics. **a.** Delineation of modeling watersheds: watersheds that are indicated in blue are drained by the Elm Fork of the Trinity River, and the watersheds indicated in red, yellow, and orange are drained by the Isle du Bois River. **b.** Physiographic regions. **c.** Land-use/land-cover classification. **d.** Soil textures.

the watershed outlet. The potential nutrient loads for total phosphorus and total nitrogen were taken from the literature and assigned to each cell based on land use, soil, and precipitation for the area. Potential soil export was estimated using the Universal Soil Loss Equation (Wischmeier and Smith 1978).

The models developed for the Texas site performed well. Total phosphorus and total nitrogen were estimated to within 5% and 15% of observed values respectively (Table 2). However, estimates of total suspended solids were only within 40% of observed values. Calculation of total loads for the entire watershed area resulted in overestimating loads by several orders of magnitude. Models consistently overestimated observed values for the Wolf Creek watershed for each parameter because of an incomplete water-quality dataset; water-quality data were only collected for the final six months in 1986.

The success in calibrating the models by increasing the stream network density demonstrated the importance of this concept for modeling hydrologically active areas and the use of the COUNT program and flow path accumulation of delivery ratios in providing this capability. Calibration steps were initiated with a COUNT threshold of 200, which approximated the blue-line stream network on the 7.5-minute USGS quadrangle maps—a standard input for many nonpoint-source models. Stream network density is related in part to soil texture and is probably not the same for different water-quality parameters, as evidenced by varying model success. (In our application the same COUNT threshold was used for all parameters within watershed for each region.)

Total suspended solids were always overestimated, and it is not necessarily valid to assume that a cell that can transport 100% of soluble nitrogen entering it can also transport 100% of the total suspended solids. Because it takes more energy to transport solids across a cell than soluble nitrogen, the stream network density for total suspended solids would logically be smaller than that for nitrogen. We also saw differences in the COUNT values to which the watersheds within different physiographic regions were calibrated. These values fell along a continuum that corresponded to the infiltration characteristics of the soils dominating each region. Thus, the clay-dominated soils in the Blackland Prairie region required a COUNT value of 5 (contributing area of 2000 m^2) to generate a stream, whereas the sandy soils of the Cross Timbers region required a COUNT value of 15 (6000 m^2; Table 2).

The application of this model to nested watersheds spanning a range of sizes (4400–100,000 ha) and having different physiographic characteristics allows us to evaluate the influence of scale on spatially dis-

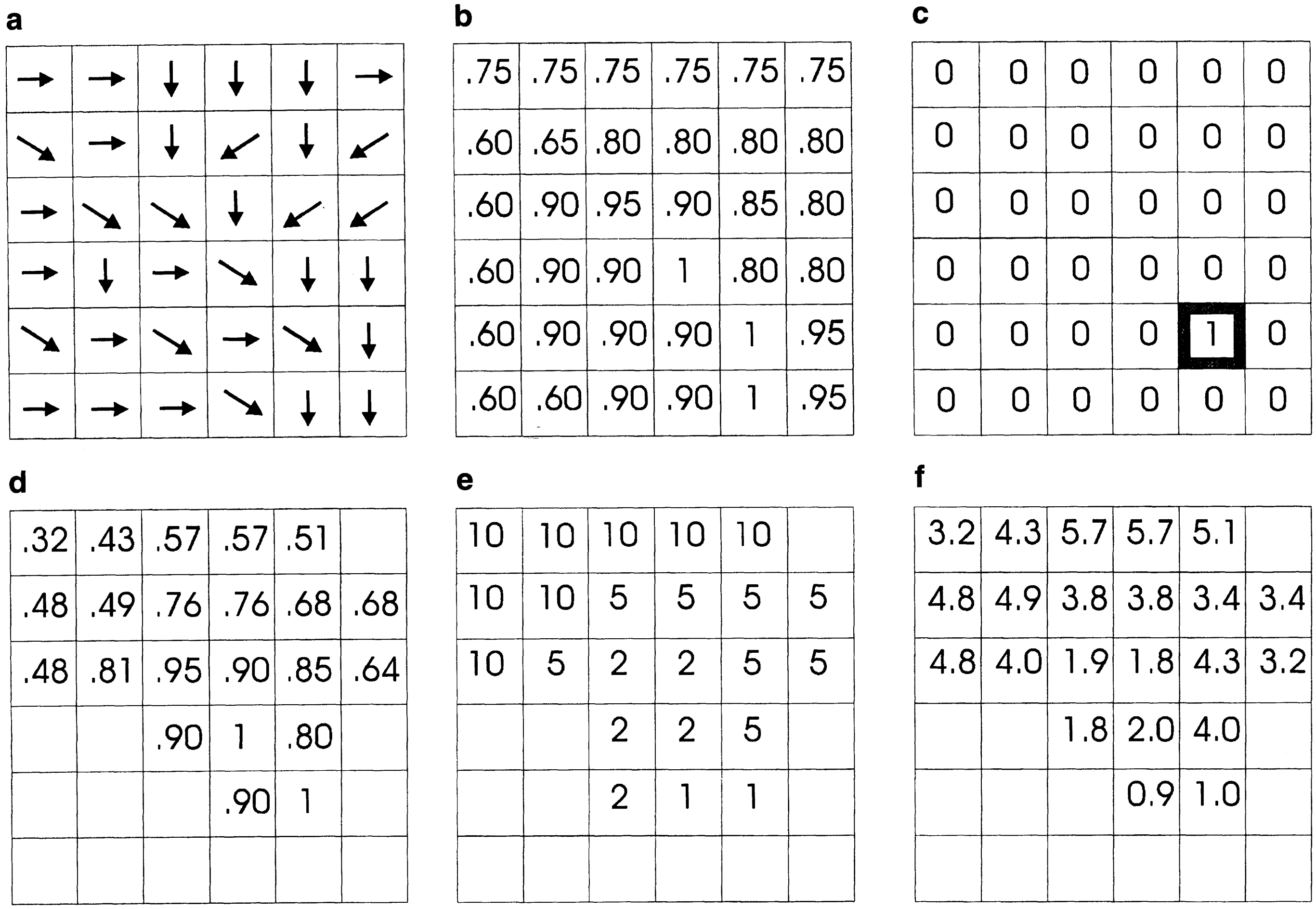

Figure 7. Geographic information system modeling procedure for the Ray Roberts study. **a.** Direction of flow from each cell is calculated based on the digital elevation model. **b.** Cell delivery ratios are calculated using regression equations and GIS layers of soil, slope, and land use. **c.** Location of watershed seed cells are based on location of water-quality sampling locations. **d.** Total flow path delivery is calculated by starting at the seed cell, or outlet, for a watershed (**c**) and using the flow direction file (**a**) to walk up the hydrologic flow path for a watershed, multiplying cell delivery ratios (**b**) together along the way. **e.** Potential nutrient loads are determined by land use and soil types using export coefficients from the literature; potential sediment load is calculated using the Universal Soil Loss Equation. **f.** Actual delivered nutrient and sediment loads from each cell are calculated by multiplying the total flow path delivery ratio file (**d**) with the potential nutrient and sediment load file (**e**).

tributed modeling. The models appear to be transportable with respect to watershed size and characteristics, although there was a slight decrease in model accuracy in the larger watersheds. The models did perform equally well in each of the physiographic regions.

Conclusions

We compared the two different studies to evaluate the use of land-use data (both the proportions and spatial pattern of land-use types) for modeling water quality and to explore how scale and data resolution influence the type of spatial analyses performed. Both studies indicated that land-use proportions are important for characterizing and modeling water quality in watersheds. However, the spatial pattern—as generalized in the Illinois study by contagion, dominance, and edges—did not greatly influence water-quality models. Additionally, the Illinois study indicates that proximity to streams is not a critical factor in modeling water quality. However, the success of the distributed model in the Texas study indicates that the location of various types of land use in the watershed is critical to modeling.

The large differences in data resolution and the fundamentally different approaches between these studies may explain these seemingly contradictory conclusions. The multivariate regression models developed for Illinois were an effort to create terrestrial metrics of water quality that would provide predictive tools for impact analysis and landscape indicators for monitoring ecological condition. Initial results (Hunsaker et al. 1992) indicated that we could accurately predict water quality using proportions of land cover and/or pattern metrics for entire watersheds (Table 1). The explanatory power of the models lessened when we used the same variables within corridors around the stream network. Omernik (1977) showed similar results by using only the proportions of forest, agriculture, and urban land uses.

For predicting annual nutrient

loadings to streams, the Texas study showed to be reasonably accurate the application of empirical equations for delivery of nutrients combined with topographic analysis to define the contributing areas. Wilkin and Jackson (1983) and Osborne and Wiley (1988) also found that land use close to the stream was a better predictor of water quality for the entire watershed than was land use.

Results from the two studies indicate that the use of the COUNT threshold method for defining a hydrologically active area may not have been appropriate at the data resolution used in the Illinois study. The COUNT threshold used in the Illinois study of 35 cells represents an area of at least 140 ha. The COUNT thresholds in the Texas study were between 8 and 15 cells, representing areas between 0.32 to 0.6 ha. Areas slightly larger than this area actually contributed nutrients and sediment to total watershed load, because the threshold technique used in the Texas study identified only areas where delivery was 100%. Nevertheless, the total upslope area contributing nutrients to any one cell was still well within the area of a single cell used in the Illinois study. This result suggests that every cell in the Illinois watersheds contributes to the water quality and explains why the models for the entire watershed were as good or better than the models using the stream-based corridors. The data resolution used in Illinois was not appropriate for modeling the hydrological active area. Because the resolution of our two datasets differed by an order of magnitude, we suggest that further work should be done with intermediate data resolutions to determine a breakpoint for the effectiveness of using hydrologically active areas.

The lumped approach employed in the Illinois study did not show any bias between watershed size and model performance, while the Texas study began to show a trend in decreasing model performance with the largest watersheds. The Texas study demonstrated that it is useful to calibrate models within similar physiographic regions (ecoregions). This finding suggests that when employing the lumped modeling approach in an area spanning a number of ecoregions, different models should be developed for each of the ecoregions.

It does not appear useful now to spend the effort (which can be substantial for large geographic areas) to create equidistance corridors or contributing areas from topography in the hope of improving lumped, empirical, nonpoint-source models using coarse-resolution data. However, if the goal is to identify areas critical for management purposes, it is important to identify hydrologically active areas and use a distributed modeling method and fine data resolution.

We conclude that management of nonpoint-source pollution in large river systems could benefit from a two-stage approach. A lumped approach with coarse-resolution data could be used as a screening method to identify watersheds making the most significant pollutant contributions. Then, a high-resolution distributed modeling technique could be used for those smaller watersheds identified as critical for specific management actions.

Acknowledgments

We wish to thank John Beauchamp and Barbara Jackson of Oak Ridge National Laboratory (ORNL) for statistical advice and computer analysis on this research. The Illinois Natural History Survey provided digital data for Illinois. Research was sponsored by the ORNL Exploratory Studies Program, US Department of Energy, and the Office of Research and Development of the US Environmental Protection Agency, under Interagency Agreement DW89934921-01 with the US Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc. Although the research described in this article has been funded in part by the US Environmental Protection Agency, it has not been subjected to agency review. Therefore, it does not necessarily reflect the views of the agency. Any mention of trade names or commercial products does not constitute an endorsement or a recommendation for use.

References cited

Adams, S. H., B. L. Kimmel, and G. R. Ploskey. 1983. Sources of organic matter for reservoir fish production: a trophic-dynamics analysis. *Can. J. Fish. Aquat. Sci.* 40: 1480–1495.

Allan, J. D., and A. S. Flecker. 1993. Biodiversity conservation in running waters. *BioScience* 43: 32–43.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper 964. US Geological Survey,Washington, DC.

Baker, W. L., and Y. Cai. 1992. The r.leprograms for multiscale analysis of landscape structure using the GRASS geographical information system. *Landscape Ecol.* 7: 291–302.

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics.* John Wiley & Sons, New York.

Cooper, J. R., J. W. Gilliam, R. B. Daniels, and W. P. Robarge. 1987. Riparian areas as filters for agricultural sediment. *Soil Sci. Soc. Am. J.* 51: 416–420.

Crowley, J. M. 1967. Biogeography. *Can. Geogr.* 11: 312–326.

DelRegno, K. J., and S. F. Atkinson. 1988. Nonpoint pollution and watershed management: A remote sensing and geographic information system (GIS) approach. *Lake Reservoir Manage.* 4: 17–25.

Environmental Systems Research Institute (ESRI). 1987. *ARC/INFO Users Guide.* ESRI, Redlands, CA.

Draper, N. R., and H. Smith Jr. 1981. *Applied Regression Analysis.* 2nd ed. John Wiley & Sons, New York.

Forman, R. T. T., and M. Godron. 1986. *Landscape Ecology.* John Wiley & Sons, New York.

Graham, R. L., C. T. Hunsaker, R. V. O'Neill, and B. L. Jackson. 1991. Ecological risk assessment at the regional scale. *Ecological Applications* 1: 196–206.

Hunsaker, C. T., R. L. Graham, G. W. Suter II, R. V. O'Neill, L. W. Barnthouse, and R. H. Gardner. 1990. Assessing ecological risk on a regional scale. *Environ. Manage.* 14: 325–332.

Hunsaker, C. T., D. A. Levine, S. P. Timmins, B. L. Jackson, and R. V. O'Neill. 1992. Landscape characterization for assessing regional water quality. Pages 997–1006 in D. H. McKenzie, D. E. Hyatt, and V. J. McDonald, eds. *Ecological Indicators.* Elsevier Applied Science, New York.

Hunsaker, C. T., R. V. O'Neill, B. L. Jackson, S. P. Timmins, D. A. Levine, and D. J. Norton. 1994. Sampling to characterize landscape pattern. *Landscape Ecol.* 9: 207–226.

Jensen, S. K., and J. O. Dominque. 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing* 54: 1593–1600.

Levine, D. A., C. T. Hunsaker, S. P. Timmins, and J. J. Beauchamp. 1993. A geographic information system approach to modeling nutrient and sediment transport. ORNL report 6736. Environmental Sci-

ences Division, Oak Ridge National Laboratory, Oak Ridge, TN.
Likens, G. E., F. H. Bormann, R. S. Pierce, J. S. Eaton, and N. M. Johnson. 1977. *Biogeochemistry of a Forested Ecosystem.* Springer-Verlag, New York.
Lotspeich, F. B. 1980. Watersheds as the basic ecosystem: This conceptual framework provides a basis for a natural classification system. *Water Resour. Bull.* 16: 581–586.
Lowrance, R., R. Todd, J. Fair Jr., O. Hendrickson Jr., R. Leonard, and L. Asmussen. 1984. Riparian forests as nutrient filters in agricultural watersheds. *BioScience* 34: 374–377.
Novotnoy, V. and G. Chesters. 1981. *Handbook of Nonpoint Pollution: Sources and Management.* Van Nos Reinhold, Atlanta, GA.
Omernik, J. M. 1977. Nonpoint source-stream nutrient level relationships: A nationwide study. EPA-600/3-77-105. US Environmental Protection Agency, Corvallis, OR.
Omernik, J. M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77: 118–125.
Omernik, J. M., A. R. Abernathy, and L. M. Male. 1981. Stream nutrient levels and proximity of agricultural and forest land to streams: Some relationships. *J. Soil Water Conserv.* 36: 227–231.
O'Neill, R. V., et al. 1988. Indices of landscape pattern. *Landscape Ecol.* 1: 153–162.
O'Neill, R. V., A. R. Johnson, and A. W. King. 1989. A hierarchical framework for the analysis of scale. *Landscape Ecol.* 3: 193–205.
Osborne, L. L., and M. J. Wiley. 1988. Empirical relationship between land use/cover and stream water quality in an agricultural watershed. *J. Environ. Manage.* 26: 9–27.
Peterjohn, W. T., and D. L. Correll. 1984. Nutrient dynamics in an agricultural watershed: Observations on the role of a riparian forest. *Ecology* 65: 1466–1475.
Pillard, D. A. 1988. Pre-impoundment estimations of nutrient loading to Ray Roberts Lake and prediction of post-inundation trophic status. Ph.D. dissertation. University of North Texas, Denton, TX.
Reckhow, K. H., M. N. Beaulac, and J. T. Simpson. 1980. Modeling phosphorus loading and lake response under uncertainty: A manual and compilation of export coefficients. EPA 440/5-80-011. US Environmental Protection Agency, Washington, DC.
Risser, P. G. 1987. Landscape ecology: State of the art. Pages 3–14 in M. G. Turner, ed. *Landscape Heterogeneity and Disturbance.* Springer-Verlag, New York.
SAS Institute, Inc. 1985. *SAS User's Guide: Statistics.* Version 5 ed. SAS Institute, Inc., Cary, NC.
Schlosser, I. J., and J. R. Karr. 1981. Water quality in agricultural watersheds: Impact of riparian vegetation during base flow. *Water Resour. Bull.* 17: 233–240.
Seaber, P. R., F. P. Kapinos, and G. L. Knapp. 1984. State hydrologic unit maps. Open File Report 84-708. US Geological Survey, Reston, VA.
Sivertun, A., L. E. Reinelt, and R. Castensson. 1988. A GIS method to aid in non-point source critical area analysis. *International Journal of Geographic Information Systems* 2: 365–378.
Timmins, S. P., and C. T. Hunsaker. In press. Tools for visualizing landscape pattern for large geographic areas. *GIS World.*
Turner, M. G., ed. 1987. *Landscape Heterogeneity and Disturbance.* Springer-Verlag, New York.
______. 1989. Landscape ecology: The effect of pattern on process. *Annu. Rev. Ecol. Syst.* 20: 171–197.
Wilkin, D. C., and R. W. Jackson. 1983. Nonpoint water quality contributions from land use. *J. Environ. Syst.* 13: 127–136.
Wischmeier, W. H., and D. D. Smith. 1978. *Predicting Rainfall Erosion Loss—A Guide to Conservation Planning.* Agriculture Handbook 537. US Department of Agriculture, Washington, DC.