SPATIALLY REFERENCED REGRESSION MODELING OF NUTRIENT LOADING IN THE CHESAPEAKE BAY WATERSHED

Stephen D. Preston, Hydrologist, U.S. Geological Survey, Baltimore, Maryland;
Richard A. Smith, Hydrologist, U.S. Geological Survey, Reston, Virginia;
Gregory E. Schwarz, Economist, U.S. Geological Survey, Reston, Virginia;
Richard B. Alexander, Hydrologist, U.S. Geological Survey, Reston, Virginia;
John W. Brakebill, Geographer, U.S. Geological Survey, Baltimore, Maryland

Abstract

A set of spatially referenced regression models is currently being developed to relate water quality in the Chesapeake Bay to sources of nutrients in the watershed and to factors that affect the transport of nutrients to the bay. Spatially referenced regression modeling is a statistical technique that uses spatial information to provide nutrient-load predictions that are more spatially detailed than those provided by other large-scale watershed models. Two applications of the technique for the determination of total nitrogen in the Chesapeake Bay watershed are described, including the estimation of incremental (local) yields and the estimation of yields delivered to the bay. The model shows that areas that are most important to the delivery of nutrients to the bay are those that drain directly to large streams or those that are near the bay. Instream loss of nutrients is minimal in both cases, thus enhancing nutrient delivery to the bay.

INTRODUCTION

Watershed modeling is commonly considered an essential tool for evaluating the sources and controls of nutrient loading to receiving waters. Watershed models provide a framework for integrating the data that describe the processes and land-surface characteristics that determine the amount of nutrients transported by streams. Development of watershed models is a difficult task, however, because of the broad spatial and temporal scales that must be considered and the large amount of information that must be integrated. Funding, time constraints and available information commonly limit the amount of spatial or temporal detail that can be considered by watershed models.

The Chesapeake Bay watershed is one area of the Nation where watershed modeling is being applied to evaluate nutrient loading (figure 1). Water quality and ecosystem integrity in the Chesapeake Bay have been affected by excessive nutrient loading, which has resulted in the depression of dissolved oxygen levels and the loss of submerged aquatic vegetation. These effects have impacted economically important aquatic species and have diminished the value of the bay as a recreational resource.

Watershed modeling has been an important component of the effort to understand nutrient loading to the Chesapeake Bay and to develop management strategies for controlling it. The Chesapeake Bay Program (CBP) is a multiagency taskforce that has been charged with coordinating and managing efforts to restore water quality in the bay. The CBP has developed a hydrologic and water-quality model for the Chesapeake Bay watershed using the Hydrologic Simulation Program - Fortran (HSPF) modeling framework (Donigian and others, 1994). Applications of the HSPF model include (1) estimating nutrient loads from all areas of the watershed, (2) evaluating the impacts of land-use-change scenarios, and (3) evaluating the potential benefits of the implementation of Best Management Practices (BMP's). The Chesapeake Bay HSPF model is temporally detailed in that it is based on hourly time increments of streamflow and other environmental processes, but is limited in spatial detail and is based on 86 segments that average more than 700 square miles in area.

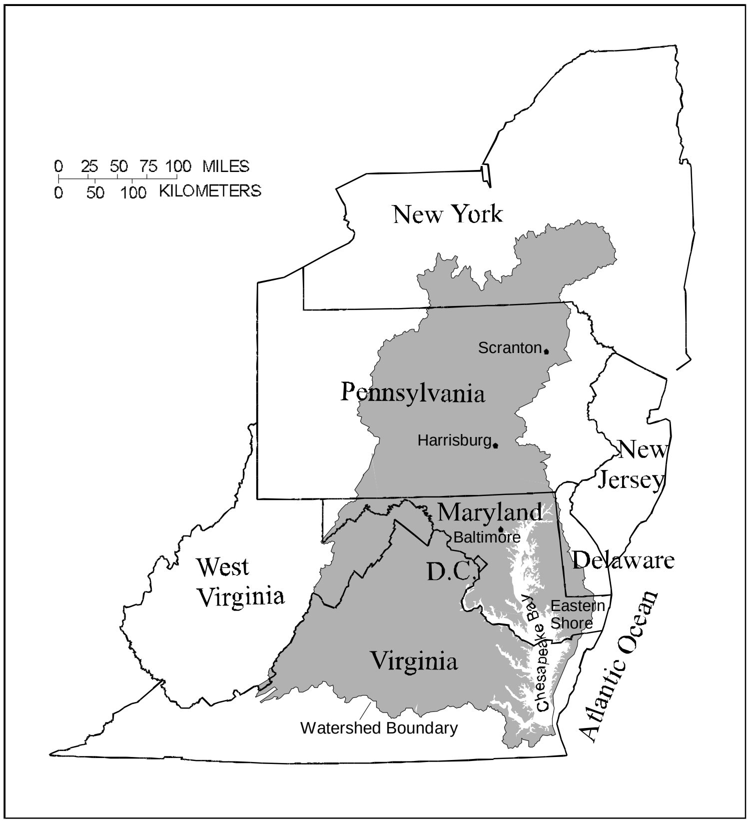

Figure 1. Chesapeake Bay watershed and surrounding area.

The HSPF modeling framework is deterministic in nature and includes a substantial amount of detail in the number of processes that are considered in simulating watershed hydrology and nutrient fate and transport. The process detail included in HSPF is important for designing and evaluating nutrient management programs. The number of parameters in the model increases with the number of processes simulated, however, and determining appropriate values for those parameters can be difficult. The current Chesapeake Bay watershed model is manually calibrated at 14 sites. Parameter values are quantified by adjusting them to fit predicted values to measured data or by adopting published values for some parameters.

To support the CBP's modeling effort, the U.S. Geological Survey (USGS) has initiated the development of a set of spatially referenced regression models. These models can be used to provide a statistical basis to watershed modeling and additional spatial detail on nutrient sources and transport processes. The method used for developing the regression models is referred to as "SPARROW" (SPAtially-Referenced Regressions On Watershed attributes) (Smith and others, 1997). The SPARROW methodology is designed to provide statistically based relations between stream-water quality and environmental factors such as contaminant sources in the watershed, land-surface characteristics that affect contaminant delivery to streams, and instream contaminant losses. Because the regression models are linked to spatial information, predictions and subsequent analytical results can be illustrated through detailed maps that provide information about nutrient loading at multiple scales. The SPARROW methodology has been successfully applied at the national scale for estimating total nitrogen and total phosphorus loads for streams in the continental United States (Smith and others, 1997).

As an initial step in the development of SPARROW models for the Chesapeake Bay watershed, this paper describes an evaluation of the national scale model within the bay watershed. SPARROW regressions are currently being developed using data that are specific to the watershed, but the national model provides a useful preliminary view of nutrient loading to the bay. Specifically, this paper describes the results of two applications of the national SPARROW model for evaluating the important sources and controls of total nitrogen loads to the Chesapeake Bay.

METHODS

The SPARROW methodology consists of a nonlinear regression in which nutrient-load data are related to upstream sources and land-surface characteristics. Spatial referencing is accomplished by linking nutrient source, land-surface characteristic, and loading information to a geographically defined river-reach data set that serves as a network for relating upstream and downstream loads. Nutrient inputs to each river reach include loading from individual sources within the watershed that drains to the reach and loading from upstream. Land-surface characteristics that affect delivery of nutrients to the reach are included by linking the relative amount of the specific characteristic in the direct drainage area to the reach. All of the dependent and independent variables are spatially defined by point or polygon coverages that are related to the stream network, which defines the connectivity and allows predictions to be presented in a spatial context. Further details of the methodology are presented below; however, the reader is referred to Smith and others (1997) and Smith and others (1993) for a complete description.

The SPARROW statistical model includes three types of parameters: source, land-to-water delivery, and instream loss parameters. The basic form of the statistical model is:

$$L_i = \sum_{n=1}^{N} \sum_{j \varepsilon J(i)} \beta_n s_{n,j} \, e^{(-\alpha' Z_j)} \, e^{(-\delta' T_{i,j})} \quad ,$$

where

L_i = load in reach *i*;

n, N = source index where *N* is the total number of considered sources;

$J(i)$ = the set of all reaches upstream and including reach *i*, except those containing or upstream of monitoring stations upstream of reach *i*;

β_n = estimated source parameter;

$s_{n,j}$ = contaminant mass from source *n* in drainage to reach *j*;

α = estimated vector of land-to-water delivery parameters;

Z_j = land-surface characteristics associated with drainage to reach *j*;

δ = estimated vector of instream loss parameters; and

$T_{i,j}$ = channel transport characteristics.

The source parameters (β_n) are included to determine the significance of individual sources in explaining the variation of loads among reaches. Sources considered in the national SPARROW model include point sources, fertilizer application rates, livestock production, atmospheric deposition and nonagricultural land. Additionally, in basins where load is monitored at some upstream location, the monitored load is considered an additional source with source parameter (β_n) set equal to one.

The land-to-water delivery parameters (α) determine the significance of different types of land-surface characteristics for increasing or decreasing the delivery of nutrients from the land surface to the stream reach. For example, relatively large percentages of impermeable surface area might be expected to increase delivery from the land surface to stream reaches. Land-surface characteristics (Z_j) that were considered in the national SPARROW model include temperature, slope, stream density, wetland, irrigated land, precipitation, and irrigated water use. Delivery of point-source loads to stream reaches was assumed to be unaffected by land-surface characteristics, and the value of the delivery term $\left(e^{(-\alpha Z_j)}\right)$ for point sources is set equal to one.

Estimation of instream loss parameters (δ) is important for relating upstream sources to downstream loads. For the national SPARROW model, instream-loss parameters were estimated for three reach classes that were defined by discharge level. The classes were defined by the discharge intervals of less than 28 m^3/s, between 28 and 283 m^3/s, and greater than 283 m^3/s.

All dependent and independent variable data sets were compiled from published data bases. Nutrient-loading data were derived from water-quality data collected as part of the USGS National Stream Quality Accounting Network (NASQAN). Load estimates were generated on the basis of total nitrogen measurements from 414 sites, including 13 sites from within the Chesapeake Bay drainage. Total nitrogen-source data were compiled primarily from published county-based data sets. Atmospheric deposition data, however, were generated through linear spatial interpolation of National Atmospheric Deposition Program (NADP) point measurements. Land surface-characteristics data were compiled from a variety of spatial data sets. Some variables were generated from county-based information (for example, wetlands, fraction of irrigated cropland). Others (soil permeability and slope), however, were compiled from the state-based soils data sets (STATSGO) (U.S. Soil Conservation Service, 1994) and published USGS data sets (temperature and precipitation). All dependent and independent variable data were compiled for calendar year 1987 or were generated to reflect conditions during that year.

The network for developing the national SPARROW model is based on River Reach File 1 (RF1) (DeWald and others, 1985) for model development and USGS hydrologic units for displaying model predictions. RF1 is a 1:500,000-scale, digital stream coverage that is attributed with reach length and average stream discharge and velocity. This information is used to classify reaches into size categories and to calculate traveltime (reach length/velocity) for estimating in-stream loss rates. Nationally, RF1 consists of approximately 60,000 stream reaches, which includes 1,366 stream reaches in the Chesapeake Bay watershed. Predicted total nitrogen loads and basin yields for the continental United States were illustrated on the basis of 2,057 USGS hydrologic "cataloging" units. For the Chesapeake Bay watershed, the scale of the basin units was refined by delineating basin boundaries for each river reach based on a 1-km^2 digital elevation model (DEM). Basin delineation produced one basin unit for each reach, or 1,366 basins in all.

Model parameters in the national model were estimated by applying a nonlinear least-squares algorithm to the equation above. The error term in the model is assumed to be multiplicative and the estimation algorithm was applied after both sides of the equation were converted to logarithmic form. The robustness of the parameter estimates was evaluated by applying a bootstrap algorithm in which the model was repeatedly estimated based on subsamples of the load and predictor data. This procedure provided distributions of model parameters that could be used to evaluate the potential range of parameter estimates. Further details and results of the bootstrap analysis are described by Smith and others (1997).

RESULTS

Regression Results and Parameter Estimates

Results of model estimation for the total nitrogen national SPARROW model are summarized in table 1. Fit of the model is good with an R-squared value of 0.87 and a mean square error of 0.4544. Most of the independent variables considered were found to be significant; variables that were clearly not significant in exploratory regressions were left out of the final model. All parametric estimates of total nitrogen source parameters were found to be significant, although livestock waste production was only moderately significant (0.0632). All bootstrap estimates of total nitrogen source parameters were found to be highly significant. Three of the eight land-to-water delivery parameters were found to be significant by the parametric or bootstrap estimations. Temperature and soil permeability were inversely related to nitrogen loading possibly because higher temperature increases rates of denitrification and because higher soil permeability tends to shift nitrate transport to ground-water reservoirs. Stream density was implemented in the model in reciprocal form and is positively related to stream nitrogen loading because basins with higher stream density are expected, on average, to have shorter overland traveltimes than basins with lower stream density. Parametric estimates of instream loss parameters were highly significant for the two smaller stream-size classes. Instream loss rates are lower for larger stream sizes because larger (deeper) streams have less contact with sediment where denitrification is expected to occur.

Application of SPARROW for Spatial Nutrient Loading Analysis

To illustrate the benefit of spatial referencing, two applications of SPARROW in the Chesapeake Bay watershed are presented (figures 2 and 3). In both cases, total nitrogen yields are calculated by dividing the predicted load by the contributing area to calculate a per unit area load. Incremental yield (figure 2) is the load generated by the area that drains directly to the reach without loads from upstream. Input to the reach is assumed to occur at the middle of the reach and instream loss is calculated over half of the length to estimate loads at the end of the reach. Incremental loads provide an indication of the relative importance of local drainage areas to nitrogen loading and provide a common basis for evaluating source areas across the entire watershed. Incremental yields provide an indication of local influences on loading, but do not account for instream losses

Table 1. Parameter estimates, probability levels and regression results for national SPARROW model (modified from Smith and others, 1997).

Model parameters	**Bootstrap Coefficient**	**Bootstrap p**
Nitrogen Sources (β)		
Point sources	0.4331	<0.005
Fertilizer application	1.439	<0.005
Livestock waste production	1.060	0.005
Atmospheric deposition	6.538	<0.005
Nonagricultural land	16.71	<0.005
Land to water delivery (α)		
Temperature	0.0198	<0.005
Slope		
Soil permeability	0.0450	<0.005
Stream density	0.0244	0.025
Wetland		
Irrigated land		
Precipitation		
Irrigated water use		
Instream loss (δ)		
δ_1 ($Q < 28.3\ m^3/s$)	0.3843	<0.005
δ_2 ($28.3\ m^3/s < Q < 283\ m^3/s$)	0.1225	<0.005
δ_3 ($Q > 283\ m^3/s$)	0.0407	0.015
R-squared	0.8742	
Mean square error	0.4544	
Number of observations	414	

that occur as nitrogen is transported to the Chesapeake Bay. The effects of trient enrichment in the bay are a jor concern and land-management agencies are seeking tools for prioritizing areas for the implementation of nutrient-reduction measures. If the bay is the primary area of concern, instream loss of nutrients is important because high local loading may become insignificant over distance and with long traveltimes. To account for instream losses, "delivered yields" (figure 3) were estimated by weighting the incremental loads by the instream loss that would occur over the distance from the end of each reach to the bay. Delivered yields provide a common basis for determining those areas in the entire watershed that are most important to the delivery of nitrogen to the bay.

Figures 2 and 3 illustrate incremental and delivered yields by shading basin areas by yield class. Areas with high incremental yields include the New York part of the watershed, southern Pennsylvania, central Maryland, western Virginia and the lower part of the eastern shore of the bay. Causes of the high local loading in these areas vary by region. Agricultural sources (fertilizer application and livestock waste production) were important to the incremental yield in most of the areas mentioned, but especially in southern Pennsylvania, central Maryland, and the Eastern Shore. In New York, agricultural sources were important, but atmospheric deposition was the primary source of nitrogen. Point sources are important in many areas of the watershed where there are large population densities, but of the areas in figure 2 with high incremental yield, point-source loading is relatively high in the Scranton, and Harrisburg, Pa., and Baltimore, Md., areas.

Comparison of figures 2 and 3 illustrates the importance of the instream loss that occurs as nitrogen is transported to the bay. Most of the areas that had relatively high incremental (local) loading were much less important with respect to loading to the bay itself. Areas with the highest delivered nitrogen loading to the bay include northeastern and southern Pennsylvania, parts of central Maryland and parts of the lower Eastern Shore. Areas with high incremental yield and relatively low delivered yield include the New York part of the watershed and parts of central Maryland and western Virginia. The highest delivered yields are areas that drain directly to large streams or are

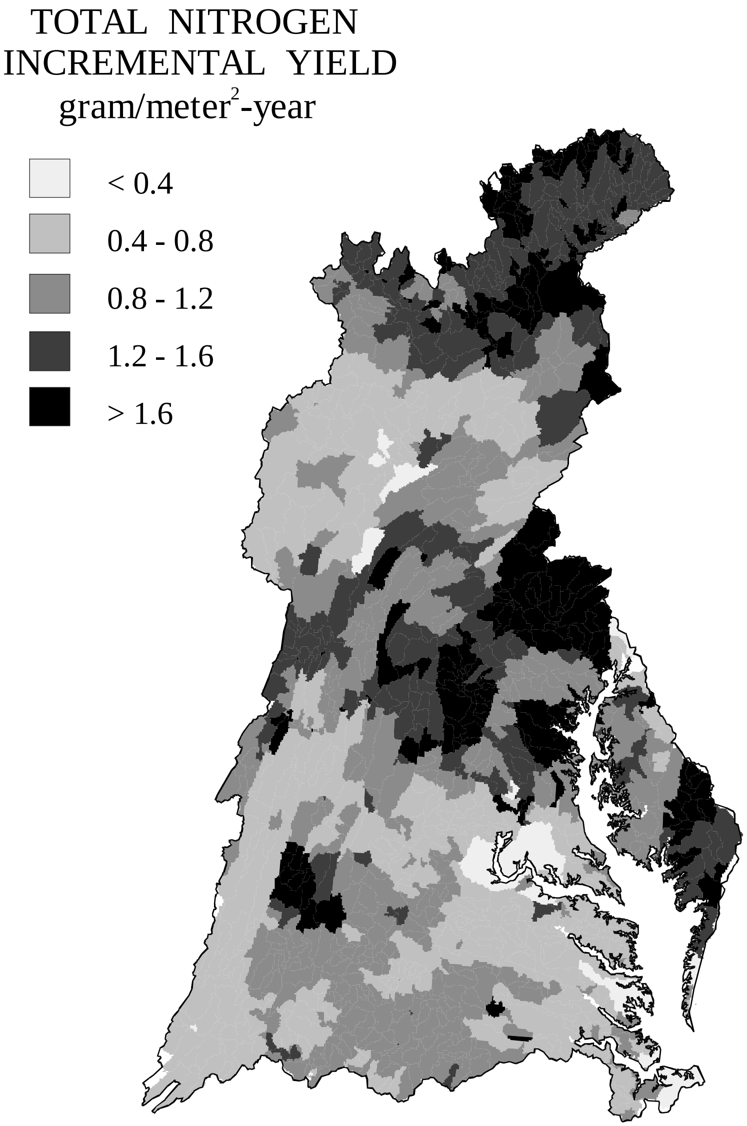

Figure 2. Incremental (local) total nitrogen yields to stream reaches in the Chesapeake Bay watershed.

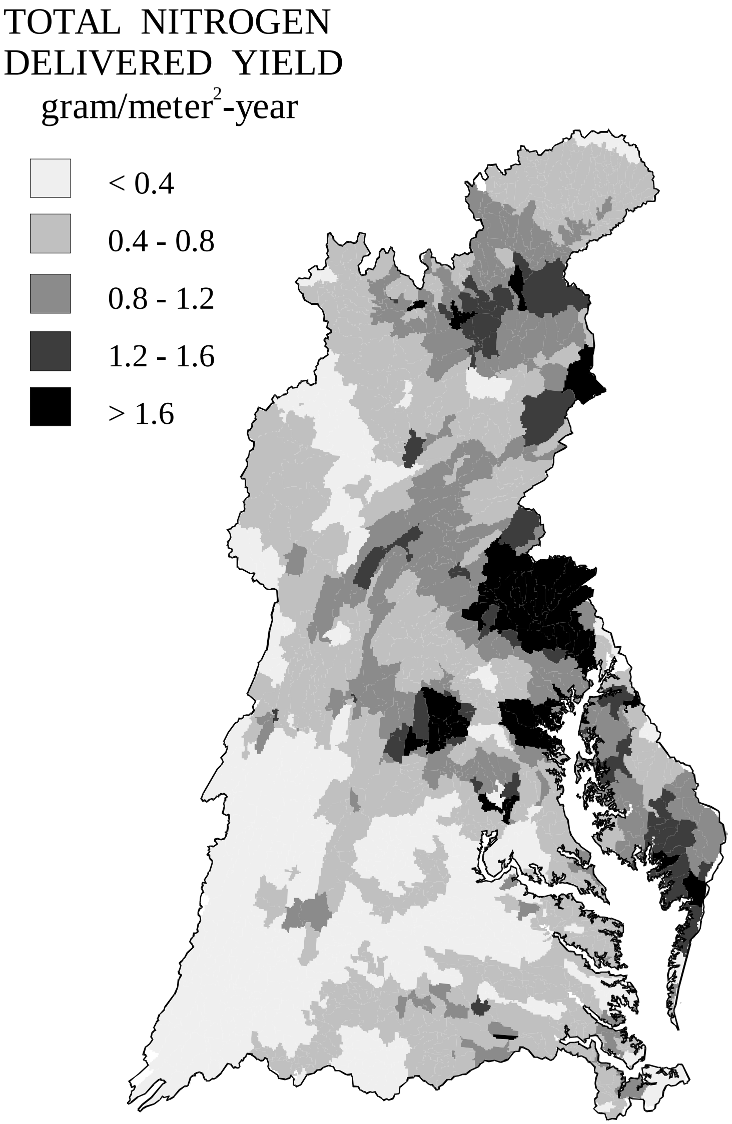

Figure 3. Total nitrogen yields delivered from stream reach drainages to the Chesapeake Bay.

areas of high incremental loading that are close to the bay. Areas that drain directly to large streams have less instream loss due to lower loss rates (table 1), and for that reason those areas may be more important for delivery of nitrogen to the bay. Areas near the bay may be more important for nitrogen delivery because travel distances are short and the time for instream loss is short compared to other parts of the watershed.

REFERENCES

Dewald, T., Horn, R., Greenspun, R., Taylor, P., Manning, L., and Montalbano, A., 1985, STORET Reach Retrieval Documentation, U.S. Environmental Protection Agency. Washington, D.C.

Donigian, A.S., Bicknell, B.R., Patwardhan, A.S., Linker, L.C., and Chang, C., 1994, Chesapeake Bay Program Watershed Model Application to Calculate Bay Nutrient Loadings -- Final Facts and Recommendations. Report # EPA 903-R-94-042, U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, Maryland, 283 p.

Smith, R.A., Alexander, R.B., Tasker, G.D., Price, C.V., Robinson, K.W., and White, D.A., 1993, Statistical Modeling of Water Quality in Regional Watersheds. Proceedings of Watershed ‘93, A National Conference on Watershed Management. Alexandria, Virginia, March 21-24, 1993, 4 p.

Smith, R.A., Schwarz, G.E., and Alexander, R.B., 1997, Regional Interpretation of Water-quality Monitoring Data. Water Resources Research, 33 (12).

U.S. Soil Conservation Service, 1994, State Soil Geographic (STATSGO) Data Base. National Soil Survey Center. Publication number 1492, 88 p.