Digital Data Used to Relate Nutrient Inputs to Water Quality in the Chesapeake Bay Watershed, Version 2.0

Open-File Report 01-251

U.S. DEPARTMENT OF THE INTERIOR
U.S. GEOLOGICAL SURVEY

Digital Data Used to Relate Nutrient Inputs to Water Quality in the Chesapeake Bay Watershed, Version 2.0

By John W. Brakebill, Stephen D. Preston, and Sarah K. Martucci

Open-File Report 01-251

Baltimore, Maryland
2001

U.S. DEPARTMENT OF THE INTERIOR
U.S. GEOLOGICAL SURVEY

U.S. Department of the Interior
GALE A. NORTON, Secretary

U.S. Geological Survey
Charles G. Groat, Director

For additional information contact:

District Chief
U.S. Geological Survey
8987 Yellow Brick Road
Baltimore, MD 21237

CONTENTS

FIGURES

Digital Data Used to Relate Nutrient Inputs to Water Quality in the Chesapeake Bay Watershed, Version 2.0

By John W. Brakebill, Stephen D. Preston, and Sarah K. Martucci

ABSTRACT

Digital data sets compiled by the U.S. Geological Survey were used as input for a collection of Spatially Referenced Regressions On Watershed (SPARROW) attributes for the Chesapeake Bay region including parts of Delaware, Maryland, New York, Pennsylvania, Virginia, West Virginia, and the District of Columbia. These regressions use a nonlinear statistical approach to relate nutrient sources and land-surface characteristics to nutrient loads of streams throughout the Chesapeake Bay watershed. A digital segmented-watershed network serves as the primary framework for spatially referencing nutrient-source and land-surface characteristic data within a geographic information system.

Flow direction and flow accumulation generated from a 30-meter cell-size Digital Elevation Model and attributes from 1:500,000-scale stream data were used to generate stream and watershed networks. Spatial data sets representing nutrient inputs of total nitrogen and total phosphorus from the early 1990's were created and compiled from numerous sources. Data include atmospheric deposition, septic systems, point-source locations, land use, land cover, and agricultural sources such as commercial fertilizer and manure. Some land-surface characteristic data sets representing factors that affect the transport of nutrients also were compiled. Data sets include land use, land cover, average-annual precipitation and temperature, slope, hydrogeomorphic regions, and soil permeability.

Nutrient-input and land-surface characteristic data sets merged with the segmented-watershed network provide the spatial detail by watershed segment required by SPARROW. Stream-nutrient load estimates for 132 sampling sites representing the early 1990's (103 for total nitrogen and 121 for total phosphorus) serve as the dependent variables for the regressions. These estimates were used to calibrate models of total nitrogen and total phosphorus depicting 1992 land-surface conditions. Examples of model predictions consist of stream-nutrient load and source percentages contributed locally to each stream reach, as well as percentages of the load that reach Chesapeake Bay.

INTRODUCTION

Nutrient enrichment and the associated hypoxia are two of the more serious problems facing the Chesapeake Bay. The U.S. Geological Survey (USGS) is investigating these issues through its Chesapeake Bay Science effort, which is designed to provide scientific information that promotes restoration and protection of the Chesapeake Bay and its watershed that include parts of Delaware, Maryland, New York, Pennsylvania, Virginia, West Virginia, and the District of Columbia. Basic goals of the USGS science effort include developing an understanding of: (1) the relation between sediment dynamics, water quality, and habitats; (2) the effect of land use and natural controls on nutrient dynamics and associated water-quality degradation; (3) the effects of increasing urbanization on water quality and habitats; and (4) the relation between living resources and associated habitats to environmental factors. A scientific understanding of the complex relations of water quality, vital habitats, and living resources within the Bay's watershed will allow resource managers to assess the effectiveness of their ecosystem restoration actions.

In 1987, the Chesapeake Bay Program (CBP), a multi-agency taskforce charged with coordinating efforts to restore water quality and ecological integrity to the Chesapeake Bay, established a restoration goal to improve dissolved oxygen levels in the bay. This goal called for the reduction of controllable nutrients entering Chesapeake Bay by 40 percent by the year 2000 (U.S. Environmental Protection Agency, 1988). Because restoration efforts fell short of achieving this goal, Chesapeake Bay has been listed as an impaired water body under the Clean Water Act and is now required to reduce dissolved oxygen levels by the year 2010. To accomplish the restoration goals, strategies from CBP management actions designed to reduce nutrient inputs entering Chesapeake Bay from its tributaries have been implemented. The USGS, an active member of the CBP, provides water-quality and living-resource information that currently is being used to evaluate and revise the standards of current nutrient-reduction strategies.

Since 1984, the CBP has been using estimates from a hydrologic and water-quality watershed model to help formulate nutrient-reduction strategies. This model, developed on the basis of a Hydrologic Simulation Program – Fortran (HSPF) modeling framework (Donigian and others, 1994), is used to estimate nutrient loads throughout the watershed. Simulations from the model are also used to evaluate land-use changes and the potential benefits of best management practices on a temporal scale.

Supporting the CBP's modeling efforts and the need for nutrient-reduction targeting and strategy evaluation, is a set of spatially referenced regression models developed by the USGS that relate nutrient sources to stream loads. SPAtially Referenced Regressions On Watershed attributes (SPARROW) is the method used to develop the regression models that retain and utilize detailed spatial information to statistically relate water-quality measurements to nutrient sources and the land-surface characteristics that affect the transport of nutrients throughout the watershed (Smith and others, 1997; Preston and others, 1998). By retaining spatial referencing, the geographical distribution and relative contribution of nutrient sources and the factors that affect nutrient transport can be examined by resource managers at various scales.

Models of total nitrogen and total phosphorus using the SPARROW methodology were successfully developed in the Chesapeake Bay watershed representing 1987 (Version I) (Preston and Brakebill, 1999; Brakebill and Preston, 1999) and currently are being completed for the early 1990's (Version II). Version II models depicting 1992 land-surface conditions, evaluate numerous spatial data sets representing potential nutrient sources for their statistical fit to the models. Spatial data sets representing manure generation, commercial fertilizer application, atmospheric deposition, land use and land cover that have shown a better statistical fit to the models are presented in this report. Data sets that were evaluated, but not used in the final models, only are referenced. Plans also are under way to develop models in the Chesapeake Bay watershed representing the late 1990's.

Purpose and Scope

This report describes digital spatial data sets generated with a geographic information system (GIS) for the purpose of applying the SPARROW methodology to develop total nitrogen and total phosphorus models in the Chesapeake Bay watershed representing the early 1990's (Version II). This report provides a description of improvements from the Version I stream reach and watershed network supporting the models. It also serves as a mechanism to distribute digital spatial data generated for and by the Version II modeling applications. Updated digital spatial data sets will be created and distributed by the USGS as planned enhancements and applications of SPARROW in the Chesapeake Bay watershed are completed.

Metadata for digital data sets related to SPARROW can be found in the appendix of this report. The metadata describes in detail the data sources, applications, methods, attributes, and procedures used to create the spatial data sets, and provides information on how to obtain the data. Spatial data discussed in this report, which are not distributed by the USGS, can be obtained by contacting agencies referenced in this report.

Acknowledgments

The authors acknowledge Robert Kellogg, U.S. Department of Agriculture, Jeffery Grimm and James Lynch, Pennsylvania State University, Robin Dennis, U.S. Environmental Protection Agency/National Oceanic Atmospheric Administration, Michael Langland, U.S. Geological Survey, and the U.S. Environmental Protection Agency's Chesapeake Bay Program Office for providing data sets for the application of SPARROW in the Chesapeake Bay watershed.

The authors also thank Richard Smith, Gregory Schwarz, and Richard Alexander of the U.S. Geological Survey, developers of the SPARROW approach, for providing technical assistance.

DATA SETS

Numerous digital data sets were created within the Chesapeake Bay watershed and surrounding area (fig. 1) from various sources using Environmental Systems Research Institute's (ESRI) Arc/Info GIS software. These data sets include an improved segmented-watershed network, water-quality data, nutrient-input sources, and land-surface characteristics. The segmented-watershed network, merged with nutrient input-source and land-characteristic data, provides input values for each stream reach to the models. The water-quality data provide information necessary for the models' calibration. All spatial data sets reside in the Albers Equal-Area Conic projection with a central meridian of 96 degrees in the North American Datum (NAD) of 1983 (Snyder, 1987). Data sets mentioned in this report are distributed for general use by the USGS, except where noted.

Water-Quality Data

A log-linear regression model called ESTIMATOR (Cohn and others, 1989) was used to estimate annual stream-nutrient loadings derived from water-quality and stream-discharge data collected by numerous State and Federal agencies (Langland and others, 1995) using the methods described in Preston and Brakebill (1999). Load-estimate regressions were calibrated based on actual stream-discharge and concentration data. Loads based on long-term average-daily discharge were estimated using the calibrated regressions, specifying the year 1992 for the trend component, and an average time series (average daily flow values for the period 1950 to 1995) for the flow terms. By use of this method, stream loads were estimated for 132 streamflow data-collection sites (103 for total nitrogen and 121 for total phosphorus) within the Chesapeake Bay watershed (fig. 2), and were used as the dependent variables (observed loads) in the calibration of SPARROW.

Many potential water-quality monitoring sites in the Chesapeake Bay watershed could not be used to estimate nutrient loads for various reasons. In some cases, sites were unusable because they did not coincide with the spatial resolution of the digital stream reach network used to spatially reference the data (described below). Many sites lacked the necessary streamflow data at or near the same location that the water-quality data were collected, or the time period of the data was not suitable to calculate loads for the early 1990's. Furthermore, water-quality sites were dropped because their representative drainage-basin size from the sampling location was inconsistent with the rest of the water-quality data set. Potential streamflow sites were not used for similar reasons.

Prior to the load calculations, latitude and longitude coordinates and station identification numbers of the data-collection sites (water quality and streamflow) were obtained from the State and Federal data bases of the various collection agencies and used to generate point-location data sets. A GIS was used to associate each water-quality sampling site with an appropriate streamflow data-collection site and stream reach. Attributes for the generated streamflow data set consist of a station identification number for the water-quality and flow-collection sites, and 1992 loading estimates for total nitrogen (103 sites) and total phosphorus (121 sites).

Segmented-Watershed Network

A segmented-watershed network (fig. 3) serves two functions. It provides the framework to spatially reference nutrient-source and land-characteristic data for input to the models, and it is used to illustrate the spatial distribution of predicted stream-nutrient loads and their potential for delivery to Chesapeake Bay. The two major components of a segmented-watershed network are stream reaches and their associated watershed boundaries. The connected stream reaches, represented by a single line that defines surface-water pathways, are consistently oriented in the direction of streamflow and are attributed with various stream characteristics used in model calibration and prediction. The previous use of the SPARROW method in the Chesapeake Bay watershed (Version I) relied on modifying an Enhanced River Reach data set (ERF1) (Alexander and others, 1999; Smith and others, 1997), and watersheds generated from coarse resolution Digital Elevation Models (DEM) for each stream reach (Brakebill and Preston, 1999). Although successful, improvements to the stream reaches and their associated watersheds (segmented-watershed network) were warranted in order to increase the level of spatial detail in areas where stream reaches were not present in the data set (coastal areas) and to improve the continuity between the stream reach locations and topographic features like elevation and slope.

Improving the existing segmented-watershed network required generating new attributed stream-reach data using available attributes from the modified ERF1 file used in Version I, and synthetic stream reaches and watershed areas generated from 30-meter digital elevation data. Seamless USGS 30-meter DEMs, organized by 4-digit hydrologic units (0205, 0206, 0207, and 0208) were acquired from a preliminary version of the National Elevation Data Base (U.S. Geological Survey, 1999a). These data contained an integer value representing elevation in meters for each cell. Flow direction, based on elevation and slope, determines the neighboring grid cell that water will flow into. A direction was calculated for each cell within each hydrologic unit. Flow accumulation, which determines the number of cells that will flow into a single cell, then was calculated for each cell within each hydrologic unit (Environmental Systems Research Institute, 1992a). A synthetic stream network was generated for each unit using a threshold of 5,000 flow-accumulated cells that will flow into a single cell. Any cell with more than 5,000 cells flowing into it represents a stream channel or water pathway. The number 5,000 was chosen as a threshold because it yielded desirable stream reaches that were comparable in scale (approximately 1:500,000) to the modified ERF1 stream reaches used in Version I. (Brakebill and Preston, 1999).

The newly generated synthetic stream reaches were converted from a raster form to a vector form and compared to the 1:100,000 scale National Hydrography Data (NHD) (U.S. Geological Survey, 1999b) for positional accuracy and density. Where the DEMs failed to yield satisfactory stream reaches (typically in flat coastal areas or near wide rivers, lakes and reservoirs) or to improve the distribution of reaches where ERF1 data are not present, the NHD vector data were inserted. The stream data by hydrologic unit then were appended, forming one seamless data set that was tested for connectivity upstream and downstream of each reach segment to ensure that proper networking topology was available (Environmental Systems Research Institute, 1992b).

The process using flow direction and flow accumulation generated more synthetic stream reaches than were necessary to build the improved segmented network. A flag in the data set was attributed if the stream reach corresponded to a modified ERF1 stream reach used in Version I. This "main channel" then was selected out of the data set to produce a subset of stream reaches that corresponded to the Version I reach data. These reaches were attributed subsequently with the same unique reach identifier that corresponds to the Version I reach file (ERF1##), creating a one-to-one relation between the two data sets. To ensure that load estimations used for model calibration were constantly referenced to the downstream end of a reach, a node was placed at each streamflow sampling location and attributed with the USGS station identification number (STAID). Nodes are endpoints of lines that maintain the identity, direction, and location of intersected linear features. This topological information is used to define reach-to-reach connectivity and allows for the identification of each reach upstream or downstream of any location along the stream network (Environmental Systems Research Institute, 1992c).

All reaches upstream of a sampling station were assigned a new unique reach identifier (E2RF1##) between 10,000 and 11,000 along with the STAID of the USGS streamflow sampling site. Placing a node at the streamflow sampling location and assigning it a unique E2RF1## value also ensured that a watershed would be generated at each calibration site on its associated reach (discussed later in this section). ERF1## and E2RF1## values are the same for any reach that is not associated with a sampling site. For any reach associated with a sampling location from Version I and not Version II, the downstream ERF1## value from Version I was calculated in E2RF1##. E2RF1## now represents the new unique stream-reach identifier for Version II and is used as a common field identifier.

Attributes essential to the modeling were transferred for each reach from the modified ERF1 data set using the reach identification number (ERF1##) as the related field in each data set. The attributes transferred included: mean water velocity, mean streamflow, hydrologic unit code, primary stream name, traveltime for reaches within reservoirs, and stream segment number. Stream-reach traveltime (reach length/velocity), an attribute necessary to estimate instream losses, was calculated separately for each reach, as was reach length. Node topology (FNODE# and TNODE#), which determines upstream and downstream connectivity of the stream reach network, also is necessary input for the models.

To improve model prediction capabilities in coastal areas, shoreline locations of major estuaries within the Chesapeake Bay watershed were added to the data set. Nodes were placed at arbitrary locations along the shoreline, creating new reach segments. Each E2RF1## was calculated for these shoreline boundaries with a unique value greater than 80,000 so that they could be easily identified. . These new segments represent the end of the transport processes to the estuaries, and only are used to spatially reference and display potential nutrient sources and predicted results.

The 30-meter flow-direction grid and the new attributed synthetic-reach network were used together to generate a 30-meter grid of watershed drainage areas for each stream reach. This process began by converting the reach network back into a 30-meter grid using the unique reach identifier (E2RF1##) as a value item. Watershed areas for each reach (all cells with the same E2RF1## value) were generated using all reach cells that represent the stream channel, or the lowest points within the watershed (Environmental Systems Research Institute, 1992a). By use of this method, all cells that represent a single reach are used as pour points, rather than only a single cell representing the absolute lowest point on the downstream end of a reach. This method also maintains the E2RF1## value from the associated reach network in the watershed data set. It serves as an identification tool as well as a common field to related data sets.

Traveltime values for shoreline watersheds that drain directly to major estuaries (E2RF1## greater than 80,000) were estimated by establishing a relation between traveltime and watershed characteristics. This relation was based on data from coastal drainages for which stream-reach traveltime was defined and watershed characteristics such as stream-reach length and mean watershed slope. In this case, the distance measured from the center of each shoreline watershed (centroid) to the nearest shoreline reach segment represents the stream-reach length used to establish the relation.

Nutrient-Input Sources

Digital data sets of atmospheric deposition, septic systems, point-source locations of nutrient discharges, land use, land cover, and agricultural fertilizer (commercial) and manure, were created and compiled from numerous sources. They represent potential nutrient-input contributions in the Chesapeake Bay watershed and were considered for the nitrogen and phosphorus models. Because numerous data sources were considered for atmospheric deposition, land use, land cover, and agricultural sources (commercial fertilizer and manure), a separate model was required for each data set in order to evaluate their statistical fit. This section describes the nutrient-input data sets and the processes used to compile and create them. Data sets are referenced that were evaluated, but not used in the final models.

Atmospheric Deposition

Three variations of atmospheric deposition data sets were evaluated for Version II SPARROW in the Chesapeake Bay watershed. They included data from the National Atmospheric Deposition Program (NADP) (National Atmospheric Deposition Program, 1993), Pennsylvania State University (Grimm and Lynch, 1997), and the Regional Atmospheric Deposition Model (National Acid Precipitation Assessment Program, 1990) for both wet and dry deposition. The NADP data represented a better statistical fit to the nitrogen models and will be discussed in this report.

Linear spatial interpolation of NADP data for 188 point measurements within the continental United States provided 1992 mean wet-deposition atmospheric estimates for nitrate (Smith and others, 1997). The latitude and longitude coordinates of the monitoring stations were used to create a spatial data set of the sampling sites attributed with the 1992 mean deposition value. A Triangulated Irregular Network (TIN) (Environmental Systems Research Institute, 1992c) was calculated by interpolating data values between the sampling locations. The area within the Chesapeake Bay region was extracted and converted into a 900-meter grid (fig. 4). This cell size was chosen because it was well suited to the spatial distribution of the monitoring stations. Smaller cell sizes did not improve the quality of the interpolated information enough to warrant such a large data set. The atmospheric-deposition grid then was used with the segmented-watershed network to calculate a mean-wet deposition value for each watershed segment.

Septic Systems

Septic system information (Fig. 5) based on 1990 census block data was obtained from NCRI Chesapeake Inc. (Maizel and others, 1995). Block group polygons were attributed with the number of persons using a septic system within that block group. To create a per capita nitrogen load value for each block group, a value of 4.24036 kilograms per person per year was multiplied by the number of persons on a septic system within each block group (Maizel and others, 1995). Each polygon then was converted to a 30-meter grid and an average nitrogen load from septic systems for each segmented watershed was calculated using the zonal functions of GRID (zonalmean) (Environmental Systems Research Institute, 1992a).

Point Sources

Locations of point-source discharges, (Fig. 6) average-annual flow, average-annual concentrations, and loads of nitrogen and phosphorus for sites within the Chesapeake Bay watershed were obtained from the CBP office for the years 1984 through 1997 (Wiedeman and Cosgrove, 1998). The data are based on information from the U.S. Environmental Protection Agency's (USEPA) Permit Compliance System (PCS) State National Pollutant Discharge Elimination System (NPDES) discharge monitoring reports, with modifications from individual State agencies responsible for monitoring point-source discharges. The original point-source data compiled by the CBP resides as monthly-discharge data, by facility, in lb/yr (pounds per year). Details on the calculation of annual loads using concentration and flow data by discharging facility can be found in Wiedeman and Cosgrove (1998).

Latitude and longitude coordinates of discharging facilities were used to generate a point data set attributed with facility name, facility type, NPDES number, and average-annual load. Average-annual load was calculated by facility for total nitrogen and total phosphorus for 1990, 1991, and 1992 and converted from (lbs/yr) to kg/yr (kilograms per year). SPARROW watershed segment numbers (E2RF1##) were assigned to the locations of each point source by merging the polygon data set representing SPARROW watershed segments with the point-source location data set. Loads of nitrogen and phosphorus for each watershed were then summed for each nutrient within each watershed segment.

Land Use and Land Cover

Land-use and land-cover data are used in various ways with SPARROW. Locations of potential nutrient sources, such as agricultural lands, are not only identified, but also are used to spatially distribute county statistics of fertilizer and manure within a watershed segment. This application serves as a means to further distribute the statistical data of a particular source used to estimate nutrient loads. Areas of potential sources and land characteristics of non-agricultural lands within each watershed segment, such as urban, forest, and wetlands, also are identified and used as inputs to the models.

Two variations of land-use and land-cover data sets were used to develop variables considered in the SPARROW models. They included Version II of the Multi-Resolution Land Characteristic (MRLC) and data obtained from the CBP that was produced by USEPA's Environmental Monitoring and Assessment Program (EMAP).

The MRLC Land Cover data[1] is a classified mosaic of 30-meter resolution Landsat 5 Thematic Mapper (TM) data generated by the multi-resolution landscape characterization project, Version II. Source TM scenes range from March of 1986 through September, 1994 with the majority of the imagery from 1991 - 1994 (Hughes-STX Corporation, 1996: Loveland and Shaw, 1996). Fifteen land-cover classifications were aggregated to seven classifications, which include open water, urban, hay/pasture/grass, row crops, forest, wetlands, and barren land (fig. 7). Acres of each land use and land cover within a SPARROW watershed segment were calculated and used as potential nutrient sources and land-surface characteristics for the models. These data was used because the time period of the source imagery was close to the desired early 1990's time period. Acres of row crop and hay/pasture within each watershed segment also were used to further distribute county manure generation statistics to calculate loading estimates of nitrogen generated from animal sources. This process is described in the Agricultural Sources section of this report.

The spatial land-cover data set provided by the CBP was derived from a combination of land-cover data sets from various years (late 1980's) including the Environmental Monitoring and Assessment Program (EMAP), National Oceanic and Atmospheric Administration (NOAA) Coastal Change Assessment Program (C-CAP), and USGS Geographic Information Retrieval and Analysis System (GIRAS) (Gutierrez-Magness and others, 1997).

[1] MRLC land cover data has been revised and updated since the completion of the 1990's SPARROW models. MRLC has developed into the National Land Cover Data

(NLCD, 2000) and as of October, 2000, can be accessed at URL http://edcwww.cr.usgs.gov/programs/lccp/mrlcreg.html. Other information regarding the MRLC consortium can be found at URL http://www.epa.gov/mrlc/data.html.

The original data set[2] was a 25-meter grid, with land-cover classifications of high-intensity urban, low-intensity urban, woody urban, herbaceous urban, herbaceous, woody, herbaceous wetland, and exposed land. Similar land-cover classifications were combined to form agriculture (herbaceous), forest (woody), urban (high-intensity, low-intensity, woody, and herbaceous urban), and wetlands (herbaceous wetland) (fig. 8). For the SPARROW applications, it was assumed that the herbaceous category within the data set represents agricultural land, and the woody category represents forest.

In order to apply the CBP land-use and land-cover spatial data, additional estimates of land use were applied. Acres of 1992 conventional-till, conservation-till, and hay land uses were calculated by the CBP within the Chesapeake Bay watershed for each county and CBP watershed model segment (CBPWS) (Donigian and others, 1994) using Crop Tillage and county Agricultural Census data bases (Gutierrez-Magness and others, 1997). To spatially distribute acres of land use within each SPARROW watershed segment, it was assumed that conventional till, conservation till, and hay land uses were distributed equally throughout the herbaceous (agricultural) classification of the CBP land-cover data set. A weighting factor was calculated by determining the percentage of CBP herbaceous (agricultural land use) land for each SPARROW watershed segment within a county and CBPWS. This calculation was done by dividing the acres of herbaceous land within a CBPWS within a county within a SPARROW watershed segment by the acres of herbaceous area within a CBPWS within a county. Acres of agricultural land use for each SPARROW watershed segment were calculated by multiplying the weighting factor and acres of each agricultural land use (conventional-till, conservation-till, and hay) within a county within a CBPWS within a SPARROW watershed segment. The resulting acres of conventional-till, conservation-till, and hay land use by CBPWS and SPARROW watershed segment were used with nutrient estimates from manure production and fertilizer application rates described in the agricultural sources section of this report (below) to calculate loading estimates of nitrogen and phosphorus from agricultural sources for each SPARROW watershed segment.

Agricultural Sources

Several data sets estimating nutrients generated from animal manure and commercial fertilizer applications were evaluated using SPARROW. Data estimating animal manure nutrients included county estimates based on the 1992 Census of Agriculture (Puckett & others, 1998), Natural Resource Conservation Service (NRCS) county estimates from all livestock and livestock held in confinement (Kellogg and others, 2000; Lander and others, 1998), and county application rates generated by the CBP (Palace and others, 1998). Fertilizer data include county estimates based on the Bureau of Census

[2] A copy of the original land-cover data set can be obtained from the Chesapeake Bay Program Office, 410 Severn Avenue, Suite 110, Annapolis, Md. 21403, (800)-968-7229.

(Lorenz, D.L., U.S. Geological Survey, written communication, 1999), and application rates generated by the CBP (Palace and others, 1998).

Using SPARROW, the total nitrogen model showed a better statistical fit using NRCS nutrient estimates recoverable from manure generated by livestock held in confinement, and fertilizer application estimates from the CBP. The total phosphorus model showed a better statistical fit using both manure and fertilizer application estimates generated by the CBP (figures 9 and 10). The process used to aggregate the data to SPARROW watershed segments is described below.

NRCS nutrient estimates from livestock manure are based on animal population numbers by county, derived from the 1992 Agricultural Census (U.S. Bureau of the Census, 1995). The basic building block of the estimation process is an animal unit, which represents 1,000 pounds of animal weight, including all beef, dairy, swine, and poultry animals. The NRCS county data included nutrient estimates from excreted and recoverable manure from all livestock and livestock held in confinement (Kellogg and others, 2000).

Several scenarios were tested to distribute the NRCS nutrient estimates throughout the row-crop and or pasture classifications (agricultural land) defined by MRLC. The scenario that best fit the models was the assumption that nutrients recoverable from manure generated by livestock in confinement would be applied to both row-crop and pasture lands (agricultural land) combined. To distribute the county nutrient estimates throughout the agricultural land within a SPARROW watershed segment, a weighting factor representing the percentage of agricultural land within each SPARROW watershed segment within a county was calculated. This calculation was accomplished by dividing the acres of agricultural land within a county and within a SPARROW watershed segment by the acres of agricultural land within a county. Load estimates by county and SPARROW watershed segment were calculated by multiplying the county NRCS nutrient estimate within a SPARROW segment by the weighting factor. Load estimates then were summed by SPARROW watershed segment to obtain load estimates of nitrogen from recoverable manure generated by confined livestock for each SPARROW watershed segment. This process distributed the county nitrogen load estimates of recoverable manure from confined livestock throughout the agricultural land within a SPARROW watershed segment.

Nutrient estimates from manure and commercial fertilizer sources were calculated by the CBP for agricultural land uses of conventional-till, conservation-till, and hay by CBPWS (Palace and others, 1998). These data differed from the NRCS data that was by weight or load, while the CBP data was by application rate, or weight of nutrient per area. It also was assumed that manure and fertilizer would be applied throughout the agricultural land as defined by the herbaceous category of the CBP land-cover data set to coincide with any other estimates created by the CBP. A weighting factor (described above in the land use section) that represents a percentage of CBP herbaceous land for each SPARROW watershed segment within a county and CBPWS was multiplied by the application rates and the total acres of conventional-till, conservation-till, and hay land uses within each CBPWS and SPARROW watershed segment. Load values for each source (manure and commercial fertilizer) by CBP agriculture land use then were summed by SPARROW watershed segment to calculate a loading estimate of total nitrogen from fertilizer and total phosphorus from manure and fertilizer for each SPARROW watershed segment. This method distributed the manure and fertilizer estimates throughout the agricultural land (herbaceous category) of each SPARROW watershed segment.

Land-Surface Characteristics

Land-surface characteristics represent potential factors that can affect the transport of nutrients and are used in SPARROW model calibrations. The sources and processes used to create and compile data sets of average-annual precipitation and temperature, slope, soil permeability, and hydrogeomorphic regions (HGMR) are described in this section. With the exception of the HGMRs and slope, the sources for land-characteristic data used for Version II of the SPARROW models are identical to the sources of land-characteristic data used in Version I (Brakebill and Preston, 1999).

Precipitation

Total monthly precipitation data and point locations for 1,695 sites within the Chesapeake Bay region were obtained from the National Climatic Data Center (NCDC) for 1950-94 (National Climatic Data Center, 1997). Average precipitation was calculated for each site and a point data set was created from the latitude and longitude coordinates provided. A TIN was calculated by interpolating data values between the sampling locations within the Chesapeake Bay region (Environmental Systems Research Institute, 1992c). The TIN was then converted to a 1-kilometer grid (fig. 11) and used with the watershed grid to calculate an average precipitation value for each SPARROW watershed segment.

Temperature

Monthly temperature data and point locations for 149 sites were obtained from the U.S. Historical Climatology Network (HCN) for 1950 - 94 (U.S. Historical Climatology Network, 1998). Average-annual temperature was calculated for each site and a point data set was created from the latitude and longitude coordinates provided. A TIN was calculated by interpolating data values between the sampling locations within the Chesapeake Bay region (Environmental Systems Research Institute, 1992c). The TIN was converted to a 1-kilometer grid (fig. 12) and used with the watershed grid to calculate an average temperature value for each SPARROW watershed segment.

Slope

Slope for the Chesapeake Bay area was calculated from a 30-meter DEM (U.S. Geological Survey, 1999a). The area containing the Chesapeake Bay watershed was extracted, and the slope function in Arc/Info's GRID module (Environmental Systems Research Institute, 1992a) was used to create a 30-meter GRID with percent slope values for each cell ranging from 0 to 125 percent (fig. 13). An average percent slope for each SPARROW watershed segment then was calculated.

Soil Permeability

Soil data originating from the State Soil Geographic Data Base (STATSGO) (Schwarz and Alexander, 1995) was converted into a 30-meter grid format using Arc/Info's GRID module. The grid was attributed with a numeric value in in/hr (inches per hour) representing the permeability of the soil. The value was calculated as a layer-thickness weighted average across soil layers of a simple average of high and low measurements of the soil layer contained in the original STATSGO data set (U.S. Soil Conservation Service, 1994). The Chesapeake Bay watershed area was extracted (figure 14), and an average permeability was calculated for each SPARROW watershed segment.

Hydrogeomorphic Regions

Hydrogeomorphic regions for the Chesapeake Bay watershed (fig. 15) were generated based on physiographic and lithologic settings (Brakebill and Kelley, 2000). Each polygon was attributed with a single HGMR code representing that region. The polygons were converted to a 30-meter grid format using Arc/Info's GRID module, and merged with the segmented watershed grid. The area of each HGMR within each SPARROW watershed segment was calculated and used as input to the models.

Nutrient Yield Estimates

Three types of nutrient estimates predicted by SPARROW for total nitrogen and total phosphorus representing 1992 land-surface conditions in the Chesapeake Bay watershed are presented in the Appendix as yields in kg/ha (kilograms per hectare) by segmented watershed. These estimates, incremental, delivered, and total, represent stream-nutrient loads from all nutrient sources evaluated by the models. Nutrient sources include point sources, septic systems, agricultural land, atmospheric deposition, and commercial fertilizer and manure application.

Incremental yield (fig. 16), which represents the local generation of nutrients, is the amount (load per area) of nutrient that is generated locally (independent of upstream load) and contributed to the downstream end of each stream reach. Each stream reach and associated watershed is treated as an independent unit, quantifying the amount of nutrient generated. Delivered yield (fig. 17) is the amount (load per area) of nutrient that is generated locally for each stream reach and weighted by the amount of in-stream loss that would occur with transport from the reach to Chesapeake Bay. The cumulative loss of nutrients from generation to delivery to the Bay is dependent on the traveltime and instream-loss rate of each individual reach. Total yield (fig. 18) is the amount (load per area) of nutrient including upstream load contributed to each stream reach. These estimates are calculated by stream reach (E2RF1##), and account for all potential sources cumulatively (Preston and Brakebill, 1999). These data, in conjunction with the segmented-watershed network, provide a useful tool for determining the spatial distribution of nutrient sources and their potential for delivery into the Chesapeake Bay.

SUMMARY

SPAtially Referenced Regressions On Watershed attributes (SPARROW) use a nonlinear statistical method to define relations among upstream nutrient-sources, downstream nutrient loads, and the land-surface characteristics that potentially affect nutrient delivery to streams. The SPARROW methodology provides a statistical basis for estimating stream-nutrient loads (predictions) and additional spatial detail on environmental factors and transport processes included in the regression models. Supporting SPARROW is a digitally based network of stream reaches and watersheds. This network provides the primary foundation for detailed spatial referencing of nutrient-source, land-surface characteristics, and nutrient predictions within a Geographic Information System. Spatial referencing provides resource managers with a tool that can be used to evaluate the geographical distribution and relative contribution of nutrient sources and the factors that affect nutrient transport at various scales.

Data derived from 30-meter Digital Elevation Models were used to generate a digital segmented-watershed network based on approximately 1,400 1:500,000-scale attributed stream reaches in the Chesapeake Bay watershed. Flow direction derived from the DEM was used to generate stream reaches and their associated watershed boundaries, creating the segmented-watershed network. Nutrient-source and land-surface characteristic data sets representing the 1992 land-surface conditions were compiled from various sources, and were merged with the segmented-watershed network to provide input information for the models. Nutrient source data sets include atmospheric deposition, septic systems, point-source locations, land use, land cover, and agricultural sources such as commercial fertilizer and manure. Land-surface characteristics include land use, land cover, average-annual precipitation and temperature, slope, hydrogeomorphic regions, and soil permeability. Nutrient-yield estimates for each segmented watershed based on model predictions are also presented in this report.

The spatial data described above are being distributed by the USGS for general use. Revised and updated digital spatial data sets will be created and distributed as planned enhancements and applications of SPARROW in the Chesapeake Bay region are completed.

SELECTED REFERENCES

Alexander, R.B., Brakebill, J.W., Brew, R.E., and Smith, R.A., Enhanced river reach file 1.2 (ERF1), 1999: U.S. Geological Survey Open-File Report 99-457 (scale 1:500,000) accessed January 10, 1999, at URL http://water.usgs.gov/GIS/metadata/usgswrd/erf1.html

Brakebill J.W. and Kelley, S.K., 2000, Hydrogeomorphic regions of the Chesapeake Bay watershed: U.S. Geological Survey Open-File Report 00-424, accessed October, 2000 at URL http://water.usgs.gov/ lookup/getspatial?HGMR

Brakebill, J.W., and Preston, S.D. 1999, Digital data used to relate nutrient inputs to water quality in the Chesapeake Bay watershed, version 1.0: U.S. Geological Survey Open-File Report 99-60, accessed May 2000 at URL http://md.usgs.gov/publications/ofr-99-60/report.html

Cohn, T.A., Delong, L.L., Gilroy, E.J., Hirsch, R.M., and Wells, D.K., 1989, Estimating constituent loads: Water Resources Research, v. 25, n. 05, p. 937-942

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.

Environmental Systems Research Institute, Inc., 1992a, Arc/Info Users Guide, Cell-based modeling with GRID, 2nd Edition: Redlands, California, 267 p.

____1992b, Arc/Info Users Guide, Network analysis, 2nd Edition: Redlands, California, 192 p.

____1992c, Arc/Info Users Guide, Surface modeling with TIN, 2nd Edition: Redlands, California, 192 p.

Grimm, J.W., and Lynch, J.A., 1997, Enhanced wet deposition estimates using modeled precipitation inputs: Final report for the U.S. Forest Service, Northeastern Forest Experiment Station, Northern Global Change Research Program: Environmental Resources Research Institute, Pennsylvania State University, 32 p.

Gutierrez-Magness, A.L., Hannawald, J.E., Linker, L.L., and Hopkins, K.J., 1997, Chesapeake Bay watershed model application and calculation of nutrient and sediment loadings, appendix E: Phase IV Watershed land use and model linkages to the airshed and estuarine models, Report # EPA 903-R-97-019: U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, Maryland, 142 p.

Hughes-STX Corporation and U.S. Geological Survey EROS Data Center, 1996 U.S. Federal Regions II and III Land Cover Data Sets, second edition.

Kellogg, R.L., Lander, C.H., Moffitt, D., and Gollehon, N., 2000, Manure nutrients relative to the capacity of cropland and pastureland to assimilate nutrients: Spatial and temporal trends for the U.S, Publication No. nps00-0579: Natural Resource Conservation Service, Washington, D.C., 93 p.

Lander, C.H., Moffitt, D., and Alt, K., 1998, Nutrients available from livestock manure relative to crop growth requirements: Resource Assessment and Strategic Planning Working Paper 98-1, 6 p.

Langland, M.J., Lietman, P.L., and Hoffman, S.A., 1995, Synthesis of nutrient and sediment data for watersheds within the Chesapeake Bay drainage basin: U.S. Geological Survey Water-Resources Investigations Report 95-4233, 121 p.

Loveland, T. R. and Shaw, D. M., 1996, Multi-resolution land characterization: Building collaborative partnerships, GAP analysis: A landscape approach to biodiversity planning. *In* Scott, J.M., Tear, T., and Davis, F., eds., Proceedings of the ASPRS/GAP Symposium, Charlotte, North Carolina, National Biological Service, Moscow, Idaho, p. 83-89.

Maizel, M., Muehlbach, G., Baynham, P., Zoerkler, J., Monds, D., Livari, T., Welle, P., Robbin, J., and Wiles, J., 1995, The potential for nutrient loadings from septic systems to ground- and surface-water resources and the Chesapeake Bay: NCRI Chesapeake Inc., based upon work supported in part by the Cooperative State Research, Education and Extension Service of the U.S. Department of Agriculture under agreement No. 95-38813-1325.

National Atmospheric Deposition Program, NADP/NTN Annual data summary: Precipitation chemistry in the United States 1993: National Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado, 1994, accessed October, 2000 at URL http://www.neonet.nl/ceos-idn/datasets/NADP_NTN.html

National Acid Precipitation Assessment Program (NAPAP), 1990, Acidic deposition: State of science and technology, Volume 1, Report 4: The regional acid deposition model and engineering model: U.S. Government Printing Office, Washington, D.C., 130 p.

National Climatic Data Center (NCDC) Monthly precipitation data for U.S. cooperative and National Weather Service sites, 1997, accessed January 12, 1998, at URL http://www.ncdc.noaa.gov/ol/ncdc.html

Palace, M.W., Hannawald, J.E., Linker, L.C., Shenk, G.W., Storrick, J.M., and Clipper, M.L., 1998, Chesapeake Bay watershed model application and calculation of nutrient and sediment loadings, appendix H: U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, Maryland, 91p. Accessed January, 2001 at URL http://www.chesapeakebay.net/pubs/777.pdf

Preston, S.D. and Brakebill, J.W., 1999, Applications of spatially referenced regression modeling for the evaluation of total nitrogen loading in the Chesapeake Bay watershed: U.S. Geological Survey Water-Resources Investigations Report 99-4054, 12 p. accessed May 2000 at URL http://md.usgs.gov/publications/wrir-99-4054

Preston, S.D., Smith, R.A., Schwarz, G.E., Alexander, R.B., and Brakebill, J.W., 1998, Spatially referenced regression modeling of nutrient loading in the Chesapeake Bay *in* Proceedings of the First Federal Interagency Hydrologic Conference: Las Vegas, Nevada, April 19-23, 1998, 8 p.

Puckett, L.J., Hitt, K.J., and Alexander, R.B., 1998, County-based estimates of nitrogen and phosphorus content of animal manure in the United States for 1982, 1987, and 1992: U.S. Geological Survey accessed May, 1999 at URL http://water.usgs.gov/ GIS/metadata/usgswrd/manure.html

Smith, R.A., Schwarz, G.E., and Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data: Water Resources Research, 33 (12) 2781 – 2798, accessed at URL http://water.usgs.gov/nawqa/sparrow/wrr97/results.html

Schwarz, G.E., and Alexander, R.B., 1995, State Soil Geographic (STATSGO) data base for the conterminous United States version 1.1: U.S. Geological Survey Open-File Report 95-449 (scale 1:250,000), accessed February 10, 1998, at URL http://water.usgs.gov/lookup/getspatial?ussoils

Snyder, J. P., 1987, Map Projections—A working manual: U.S. Geological Survey Professional Paper 1395, 383 p.

U.S. Bureau of the Census, 1995, 1992 Census of Agriculture: The Bureau, Washington D.C. (machine readable data files).

U.S. Environmental Protection Agency, 1988, Chesapeake Bay Agreement, *in* Comprehensive research plan, an agreement commitment report from the Chesapeake Executive Council [variously paged].

U.S. Geological Survey, 1999a, USGS National Elevation Data Base (NED): U.S. Geological Survey Fact Sheet 148-99, 1 p. accessed May 2000 at URL http://mapping.usgs.gov/mac/isb/pubs/factsheets/fs14899.html

____1999b, USGS National Hydrography Data Base (NHD): U.S. Geological Survey Fact Sheet 106-99, 1 p. accessed May 2000 at URL http://mapping.usgs.gov/mac/isb/pubs/factsheets/fs10699.html

____2000, National Land Cover Data Base (NLCD), accessed October 2000 at URL http://edcwww.cr.usgs.gov/programs/lccp/mrlcreg.html.

U.S. Historical Climatology Network (HCN) Serial Temperature Data, accessed January 1998, at URL http://www.ncdc.noaa.gov/ol/climate/research/ushcn/ushcn.html

U.S. Soil Conservation Service, 1994, State Soil Geographic (STATSCO) data base: National Soil Survey Center, Washington, D.C., Publication number 1492, 88 p.

Wiedeman, A., and Cosgrove, A., 1998, Chesapeake Bay watershed model application and calculation of nutrient and sediment loadings, appendix F: Report # EPA 903-, U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, Maryland, 50 p.

APPENDIX

Metadata for Digital Data Sets