A HYDROLOGIC NETWORK SUPPORTING SPATIALLY REFERENCED REGRESSION MODELING IN THE CHESAPEAKE BAY WATERSHED

JOHN W. BRAKEBILL[1*] AND STEPHEN D. PRESTON[2]

[1]U.S. Geological Survey, Baltimore, MD, USA;
[2]U.S. Geological Survey, Chesapeake Bay Program, Annapolis, MD, USA
[*](author for correspondence, email; jwbrakeb@usgs.gov)

Abstract. The U.S. Geological Survey has developed a methodology for statistically relating nutrient sources and land-surface characteristics to nutrient loads of streams. The methodology is referred to as SPAtially Referenced Regressions On Watershed attributes (SPARROW), and relates measured stream nutrient loads to nutrient sources using nonlinear statistical regression models. A spatially detailed digital hydrologic network of stream reaches, stream-reach characteristics such as mean streamflow, water velocity, reach length, and travel time, and their associated watersheds supports the regression models. This network serves as the primary framework for spatially referencing potential nutrient source information such as atmospheric deposition, septic systems, point-sources, land use, land cover, and agricultural sources and land-surface characteristics such as land use, land cover, average-annual precipitation and temperature, slope, and soil permeability. In the Chesapeake Bay watershed that covers parts of Delaware, Maryland, Pennsylvania, New York, Virginia, West Virginia, and Washington D.C., SPARROW was used to generate models estimating loads of total nitrogen and total phosphorus representing 1987 and 1992 land-surface conditions. The 1987 models used a hydrologic network derived from an enhanced version of the U.S. Environmental Protection Agency's digital River Reach File, and course resolution Digital Elevation Models (DEMs). A new hydrologic network was created to support the 1992 models by generating stream reaches representing surface-water pathways defined by flow direction and flow accumulation algorithms from higher resolution DEMs. On a reach-by-reach basis, stream reach characteristics essential to the modeling were transferred to the newly generated pathways or reaches from the enhanced River Reach File used to support the 1987 models. To complete the new network, watersheds for each reach were generated using the direction of surface-water flow derived from the DEMs. This network improves upon existing digital stream data by increasing the level of spatial detail and providing consistency between the reach locations and topography. The hydrologic network also aids in illustrating the spatial patterns of predicted nutrient loads and sources contributed locally to each stream, and the percentages of nutrient load that reach Chesapeake Bay.

Keywords: SPARROW, Chesapeake Bay, nutrients, loads, regression modeling, reach, stream, RF1, DEM

1. Introduction

SPAtially Referenced Regressions On Watershed attributes (SPARROW) use a nonlinear statistical method that defines 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) as well as additional spatial detail on environmental factors and transport processes included in the regression models. This method, developed by the U.S. Geological Survey (USGS) at a national scale, addresses data-interpretation limitations of regional water-quality assessments caused by sparse sampling, sampling bias, and basin heterogeneity (Smith *et al.*, 1997). In the Chesapeake Bay watershed (Figure 1), a 165,000 km^2 (square-kilometer) watershed that covers parts of Delaware, Maryland, New York, Pennsylvania, Virginia, West Virginia, and Washington, D.C., the regression models are used to explore the application at regional scales, while providing resource managers with useful information about nutrients entering Chesapeake Bay and its tributaries (Preston and Brakebill, 1999).

SPARROW models estimating nutrient loads of total nitrogen and total phosphorus were developed in the Chesapeake Bay watershed representing 1987 and 1992 land-surface conditions. These models estimate the amount of total nitrogen and total phosphorus generated locally in tributary watersheds, as well as how much of the nitrogen and phosphorus generated locally reaches the Chesapeake Bay, taking into account the amount of in-stream decay that may occur during transit (Preston and Brakebill, 1999).

Supporting the models is a geographically referenced hydrologic network of connected streams and associated watersheds. The stream network is based on an enhancement of the U.S. Environmental Protection Agency (USEPA) digital River Reach File (RF1) (U.S. Environmental Protection Agency; 1996; Alexander *et al.*, 1999). RF1, a vector database based on 1:500,000-scale mapping containing locations of stream reaches and associated characteristics, defines a stream reach with a single line representing surface-water pathways that extends either from headwater to stream junction, or from one stream junction to another stream junction. Enhancements to the USEPA RF1 database (ERF1) improved stream-reach characteristics such as mean water velocity, streamflow, time of travel, and reservoir information that are necessary to calculate stream-channel transport parameters for the SPARROW models (Smith *et al.*, 1997; Alexander *et al.*, 1999).

Watershed boundaries for each stream-reach provide the basis for spatially referencing potential nutrient sources, land-surface characteristics, and nutrient predictions. Atmospheric deposition, septic systems, point-sources, land use, land cover, and agricultural sources such as applied commercial fertilizer and manure are all potential nutrient sources examined by SPARROW. Land use, land cover, average-annual precipitation and temperature, slope, and soil permeability are land-surface characteristics that potentially affect nutrient transport and delivery. By retaining spatial referencing, the geographical distribution and relative contribution of nutrient sources and the factors that affect nutrient transport can be examined at various scales (Preston and Brakebill, 1999; Brakebill and Preston, 1999).

The 1:500,000-scale ERF1 data and watersheds generated from derivative products of a 1-km (kilometer) cell Digital Elevation Model (DEM) (U.S. Geological Survey, 1997) provided the basis for the network supporting the 1987 models. These data presented limitations in spatial location and detail, but possessed stream-reach characteristics necessary for the models. The limitations in the characterized stream-reach data and higher resolution elevation data on a regional scale

Figure 1. Location of the Chesapeake Bay watershed.

limited the accuracy and precision of the 1987 network. Improvements to stream locations and their associated watersheds were warranted in order to increase the level of spatial detail in areas where stream reaches were not present in the RF1 data set (coastal areas) and to improve the continuity between the stream-reach locations and topographic features such as elevation and slope.

The 1992 models introduced improvements to the network by applying more detailed DEM data to recreate a stream-reach network and its associated watersheds. Stream-reach characteristics (mean streamflow, velocity, and time of travel) from the enhanced RF1 (ERF1) data and seamless 30-m (meter) DEMs are key elements of the 1992 network's improved utility (Brakebill *et al*., 2001). This paper briefly describes the functionality of a hydrologic network supporting spatially referenced regression modeling in the Chesapeake Bay watershed (Section 2). The methods used to improve the network and examine the implications for SPARROW model development are described in Section 3.

2. Network Description

Total nitrogen and total phosphorus SPARROW models applied in the Chesapeake Bay watershed utilize a connected hydrologic network of stream reaches and associated characteristics and watersheds to geographically reference predicted and observed stream-nutrient loads, potential nutrient sources, and land-surface characteristics to individual stream reaches. Nitrogen and phosphorus sources associated with an individual stream reach, the relative contribution of each source to downstream loads, and the relation between the model's nutrient load predictions can all be spatially evaluated since all of the information is referenced to the network. Annual stream nutrient-loading estimates, referred to as "observed", are derived from water-quality and stream-discharge data collected by numerous State and Federal agencies. The data provide the downstream loading information and serve as the dependent variable in the calibration of the SPARROW models (Preston and Brakebill, 1999; Langland *et al*., 1995).

Stream-reach data are the foundation for spatially referencing nutrient load (observed), source, and transport data, and SPARROW predictions of nutrient loads. Based on the Enhanced River Reach File (ERF1), an enhancement of the 1:500,000-scale USEPA RF1, the reach data contains spatially networked topological properties and stream-reach characteristics (attributes) essential to the model's calibration and prediction capabilities (Alexander *et al*., 1999). Every stream reach is consistently oriented in the direction of streamflow and is connected to at least one other reach at its downstream node. 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 (ESRI), 1992a]. Linear connection is important because it allows for systematic climbing upstream from each monitoring station, aggregating the observed load to every reach upstream to the next monitoring station. This aggregation ensures that each reach upstream of a monitoring station receives representation of the observed load from the downstream location. In the model prediction phase, the design of the network also permits the accumulation of nutrients from each reach as they move downstream toward Chesapeake Bay.

Important stream-reach characteristics include estimates of mean streamflow and water velocity, reach length and travel time (calculated as a ratio of reach length over mean stream-water velocity), and a unique reach identification number. Because SPARROW statistically relates upstream nutrient sources to observed downstream loads, the methodology takes into account nutrient losses due to natural in-stream processes such as sedimentation and denitrification. In-stream nutrient-loss rates, which can vary by stream size, are estimated statistically for stream classes using the mean annual streamflow and travel time values from the stream-reach data (Smith *et al.*, 1997). The unique identification number is used to associate all load, source, transport, and prediction information to each reach. A unique number identifies each reach as a single unit, and allows the aggregation of nutrient source, transport, and prediction data to these units.

Associated watershed boundaries generated from coarse resolution DEMs for each stream reach provides the spatial detail of the 1987 SPARROW models (Brakebill and Preston, 1999). Whereas the stream-reach network provides the linear connection of surface-water pathways, watersheds attributed with a unique identification number of their associated reach provide the ability to spatially reference nutrient source, load, transport, and load prediction data for each reach on an area basis. Nutrient-source data merged with the watersheds produce load values for each reach watershed and are used as input values for the models. Land-surface characteristics also are merged with the watersheds to produce average watershed characteristics for each stream reach. Nutrient sources, and predicted nutrient loads displayed by individual watersheds as yields, are evaluated for spatial distribution, relative importance, and potential for delivery to Chesapeake Bay.

3. Network Improvements

Limitations in the ERF1 stream-reach data and the coarse resolution of the DEMs used to generate associated watersheds from the 1987 network necessitated further evaluation. The ERF1 stream reaches poorly represent the actual locations of stream channels when compared to other topographic features such as elevation and slope. Stream-reach densities varied throughout the Chesapeake Bay watershed due to different mapping procedures used to compile the data. Coastal area streams along major estuaries are not represented; in the Potomac River estuary, only major tributaries draining directly into the estuary are represented in the

ERF1 data set. This representation could exclude potential nutrient point sources such as sewage treatment plants within the current network. Single lines in the center of major estuaries are present in the original RF1 and were not used because they represent tidal processes that were too complex to incorporate into the calibration phase of the models. Errors in the stream-network connectivity and the accuracy in stream locations at confluences also were a concern.

When the 1987 models were developed, the 1-km cell DEM acquired to generate watershed boundaries for each reach was the only consistent source of topographic information available on a regional scale (U.S. Geological Survey, 1997). Other elevation data at various resolutions were considered, but required extensive processing time or cost for compilation.

In order to improve the network for the 1992 Chesapeake Bay SPARROW models, a method was developed that increased the level of spatial detail in areas where stream reaches were not present while improving the continuity among the stream-reach locations, watershed boundaries, and topographic features. This method required the construction of synthetic stream reaches and watershed areas generated using seamless DEMs from the USGS National Elevation Data (NED) (U.S. Geological Survey, 1999a), and the transfer of existing stream-reach characteristics from the ERF1 file (Brakebill *et al.*, 2001).

3.1 STREAM REACHES

Improvements to the stream-reach network utilized seamless 30-m cell DEMs to generate the water pathways that would make up the stream-reach network (U.S. Geological Survey, 1999a). Components key to constructing water pathways are the direction of surface-water flow from each elevation cell (flow direction), and the accumulation of cells flowing into any given cell (flow accumulation). Flow direction represents the steepest downslope direction water on a surface will flow. Once this direction is known, the identification and the number of cells flowing into a given cell can be calculated, and used to generate stream networks and watershed boundaries (Environmental Systems Research Institute, 1992b).

Flow direction from each 30-m elevation cell within the Chesapeake Bay watershed was calculated to one of its eight adjacent or diagonal neighboring cells. Flow accumulation was calculated for each 30-m cell using the flow-direction information. Based on the flow-accumulated cells, a streamwater pathway was generated by applying a threshold of 5,000 cells that will flow into any single cell. Any cell with more than 5,000 cells flowing into it represents a water pathway, or stream reach. The number 5,000 was chosen as a threshold because it yielded desirable pathways that were comparable in scale to the ERF1 stream reaches. This threshold was necessary so that attributes from the ERF1 data could be used with the newly created stream reaches, simplifying the transfer of attributes from ERF1.

The newly generated stream reaches were converted from a raster form to a vector form. Because the synthetic reaches were generated using elevation information, this process maintained the directional topology of the reaches, such that the direction of the topology was consistent with the direction of streamflow, allowing for the movement upstream and downstream of any given reach. Positional accuracy and density of the reach data were evaluated by comparing them to the USGS National Hydrography Dataset (NHD) 1:100,000-scale stream data (U.S. Geological Survey, 1999b). 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 did not exist, the NHD vector data were inserted.

The synthetic reach generation process created more water pathways than were necessary to build the improved network. Stream reaches corresponding to the scale of the ERF1 stream-reach data were identified and selected out of the data set to produce a subset of generated synthetic stream reaches. Each new reach was attributed with the same unique reach identifier that corresponds to the ERF1 reach, creating a one-to-one relation between the two data sets. This relation provided the means to transfer stream-reach characteristics from one data set to another. A comparison of location among ERF1, topography, and the newly generated stream reaches using DEMs is shown in Figure 2.

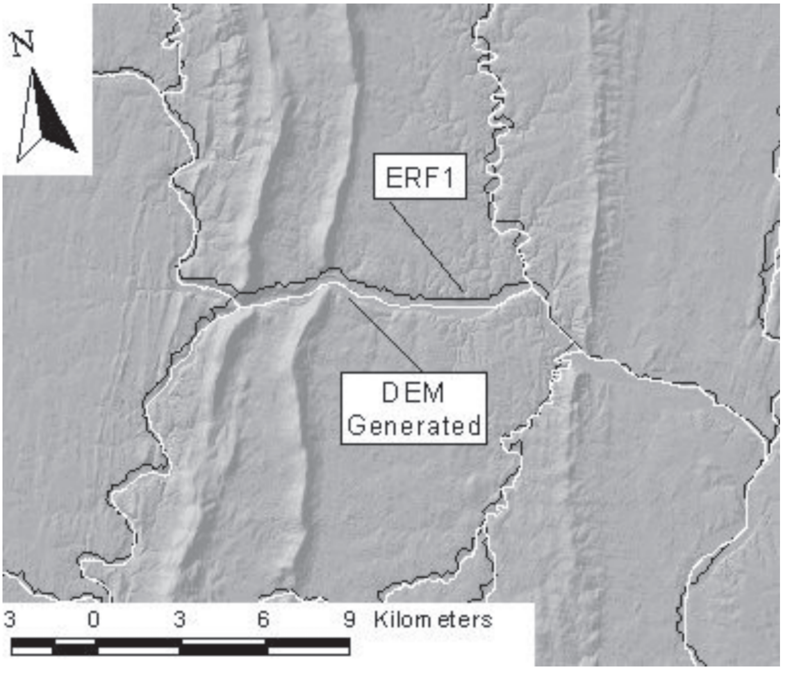

Figure 2. Spatial relation among newly generated stream-reaches using 30-m DEM (Digital Elevation Models), ERFI reaches, and topographic features.

To ensure that load estimates used for model calibration were referenced to the downstream end of a reach, a node was placed at each streamflow sampling location. By adding a node at the sampling location, the reach was segmented into two separate stream reaches, creating a new reach upstream of the sampling site and maintaining the reach connectivity. This segmentation ensured that a watershed from each sampling site location along its associated reach would be generated, allowing for the distribution of observed loads upstream of the sampling sites.

3.2 WATERSHEDS

Watersheds are important to the SPARROW models because they provide the ability to reference nutrient sources and predicted results in spatial detail. To improve the resolution in coastal areas for spatial referencing, nodes were placed at arbitrary locations along the estuary shoreline boundaries, creating new reach segments. This improved resolution ensures that during the watershed generation process, areas draining directly to the coastal estuaries would be included in the network. 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 (Figure 3).

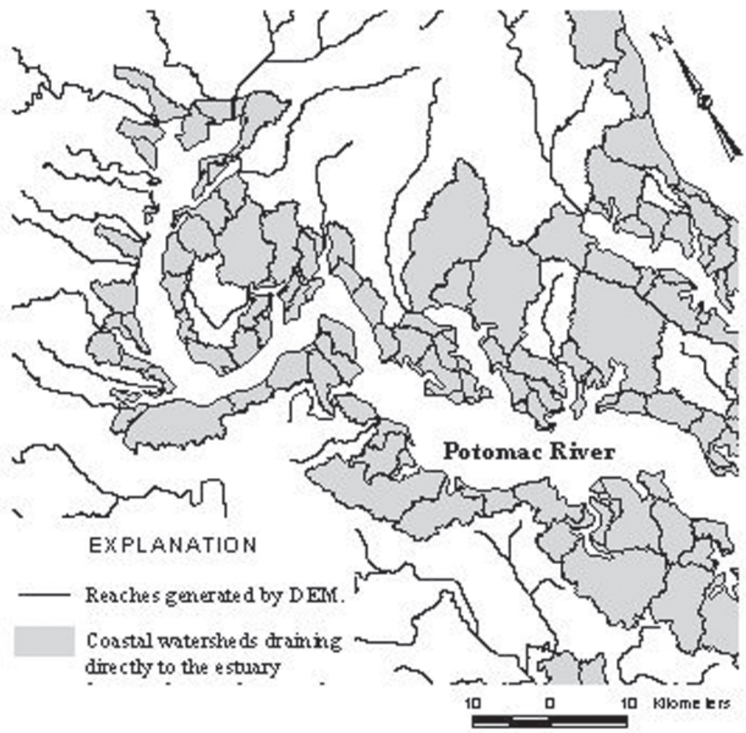

Figure 3. Coastal watersheds that drain directly into the Potomac River estuary.

The 30-m flow-direction data and the new attributed synthetic-reach network were used to generate 2,249 watershed drainage areas for each stream reach and segmented shoreline boundary. This result was accomplished by converting the reach network back into a 30-m grid using the unique reach identifier number to populate multiple cells that represent a stream reach with the same corresponding unique value. Watershed areas for each reach were generated using all reach cells, which represent the water pathway, or the lowest points within the watershed (Environmental Systems Research Institute, 1992b). In this method, all cells that represent a single reach are used as locations for the watershed generation process to begin, instead of a single cell representing the absolute lowest point on the downstream end of a reach. This method also maintains the unique reach value from the associated reach network in the watershed data set, serving as an identification tool as well as a common field to related data sets.

The watersheds generated for the 1992 SPARROW models represent a vast improvement from the 1987 network. Finer resolution elevation data improved the locations of the boundaries with respect to the topography, and eliminated noticeable boundary generalizations (Figure 4).

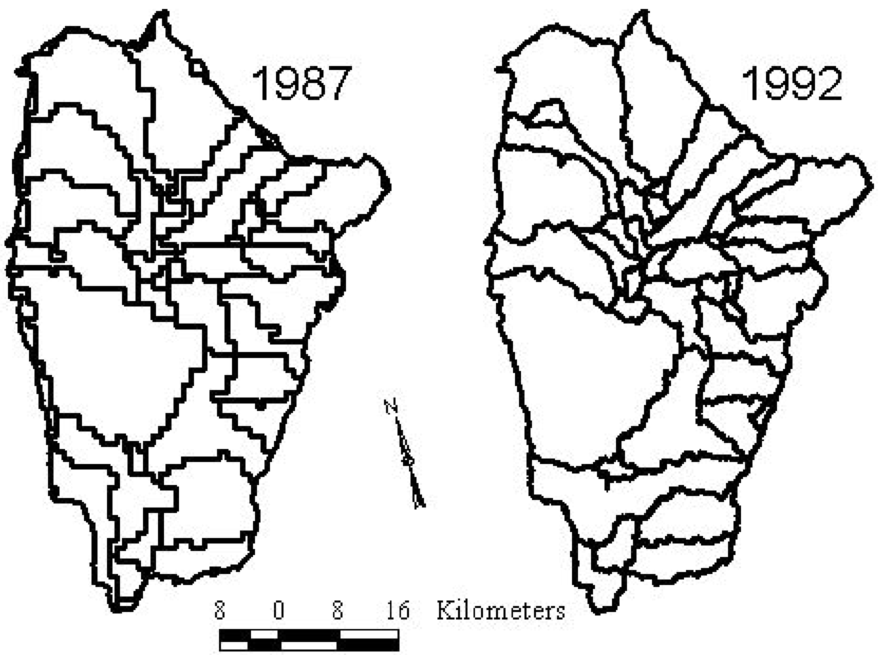

Figure 4. Comparison watersheds of the 1987 network based on a 1-km DEM (Digital Elevation Model) and the 1992 network based on a 30-m DEM.

3.3 ATTRIBUTED NETWORK

ERF1 contains stream-reach characteristics essential to the SPARROW modeling applications. This information is necessary to calculate stream-channel transport parameters used in by models. ERF1 is the main component of the hydrologic network supporting SPARROW, and will continue to be a valuable asset to modeling networks until better methods for estimating necessary stream characteristics are developed (Smith et al, 1997; Alexander *et al.*, 1999).

Each newly generated stream reach was populated with ERF1 stream characteristics, creating an attributed stream network. This process utilized the unique reach identification number in both data sets as a common field. Characteristics necessary to the model's functionality then were transferred from ERF1 to the new stream-reach data set. These characteristics include mean water velocity, mean streamflow, and travel time for reaches within reservoirs. Travel time for stream reaches not within reservoirs was calculated separately for each reach, as was reach length. This calculation included reaches with nodes placed at sampling site locations. Once the attributes were transferred and travel times calculated, the new reaches upstream of the sampling sites were issued a new unique identification number and attributed with the sampling site identification number of their immediate downstream site.

Travel time values for shoreline watersheds that drain directly to major estuaries were estimated by establishing a relation between travel time and watershed characteristics (Figure 5). This relation was based on data from coastal drainages for which stream-reach travel time 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 to the nearest shoreline reach segment represents the stream-reach length used to establish the relation. The relation based on existing travel time data provided a reasonable R^2 (0.71) and is shown in Figure 5. In general, watersheds with greater slopes and shorter stream-reach lengths were assigned shorter travel times.

4. Summary

Spatially referenced regression modeling utilizes a nonlinear statistical approach for relating upstream nutrient sources to downstream nutrient loads. Referred to as SPARROW, the methodology was used to generate models estimating loads of total nitrogen and total phosphorus representing 1987 and 1992 land-surface conditions in the Chesapeake Bay watershed. Supporting SPARROW is a hydrologic network of connected stream reaches and watersheds that provide spatial referencing of nutrient sources, observed and predicted stream-nutrient loads, and land-surface characteristics.

A network supporting the 1987 models was based on stream locations and characteristics from an enhanced version of the USEPA's River Reach File (ERF1) and watersheds generated using a 1-km cell DEM. Although successful, this process produced a network containing general reach locations and watershed boundaries. A new method was developed to generate a hydrologic network supporting the 1992 models that used seamless 30-m DEMs to generate water pathways representing stream-reach locations and their associated watersheds. This result produced spatial improvements from the 1987 network including water pathways and their associated watershed boundary locations matching topographic features. Stream-reach characteristics necessary for SPARROW's functionality such as mean streamflow and travel time were transferred from the ERF1 data set used in the 1987 network to the newly generated stream reaches. Coastal areas of the network were improved spatially by adding reaches to the data set, generating watersheds for areas that drain directly to coastal estuaries, and estimating necessary travel time information for these areas based on stream reaches with similar watershed characteristics.

$$\text{Travel Time} = \beta_0 + \beta_1 \text{ Centroid} + \beta_2 \text{ Slope}$$

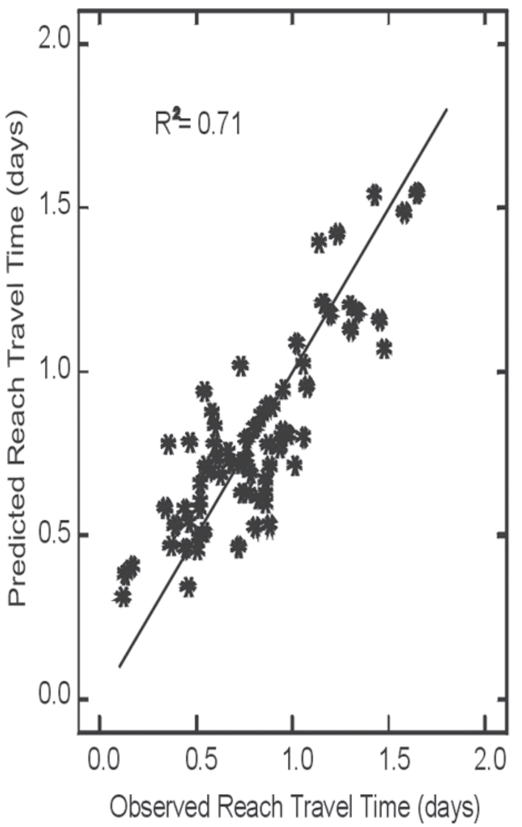

Figure 5. Relation of predicted and observed stream-reach travel time.

5. References

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

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

Brakebill, J.W., Preston, S.D., and Martucci, S.K.: 2001, Digital Data Used to Relate Nutrient Inputs to Water Quality in the Chesapeake Bay Watershed, Version 2.0: U.S. Geological Survey Open-File Report 01-251, accessed July 2001 at URL http://md.usgs.gov/publications/ofr-01-251/report.html

ESRI (Environmental Systems Research Institute, Inc): 1992a, Arc/Info Users Guide, ARC/INFO Data Model, Concepts and Key Terms, 2d ed.: Redlands, California, 201 p.

ESRI (Environmental Systems Research Institute, Inc): 1992b, Arc/Info Users Guide, Cell-based modeling with GRID, 2d ed.: Redlands, California, 267 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.

Preston, S.D. and Brakebill, J.W.: 1999, Application 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. Available online at URL http://md.usgs.gov/publications/wrir-99-4054/report.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, available online at URL http://water.usgs.gov/nawqa/sparrow/wrr97/results.html

USEPA (U.S. Environmental Protection Agency): 1996, Reach File Version 1.0 (RF1) for the conterminous United States: U.S. Environmental Protection Agency accessed January 1998 at http://www.epa.gov/ost/basins/metadata/rf1.htm

USGS (U.S. Geological Survey): 1997, HYDRO1K Elevation Derivative Data Base: U.S. Geological Survey accessed January 1998 at URL http://edcwww.cr.usgs.gov/landdaac/gtopo30/hydro/namerica.html

USGS (U.S. Geological Survey): 1999a, National Elevation Data Base (NED): U.S. Geological Survey Fact Sheet 148–99, available online at URL http://mapping.usgs.gov/mac/isb/pubs/factsheets/fs14899.html

USGS (U.S. Geological Survey): 1999b, National Hydrography Data Base (NHD): U.S. Geological Survey Fact Sheet 106–99, available online at URL http://mapping.usgs.gov/mac/isb/pubs/factsheets/fs14899.html http://mapping.usgs.gov/mac/isb/pubs/factsheets/fs10699.html